THE FACTORIAL ANALYSIS
OF HUMAN ABILITY

THE FACTORIAL ANALYSIS OF HUMAN ABILITY

By

GODFREY H. THOMSON, D.Sc., Ph.D.

Professor of Education in the University of Edinburgh

UNIVERSITY OF LONDON PRESS LTD.

10 & 11 WARWICK LANE, LONDON, E.C.4

First Published . . . *January* 1939

AGENTS OVERSEAS

AUSTRALIA, NEW ZEALAND
AND SOUTH SEA ISLANDS
W. S. SMART, P.O. Box 120 C C,
SYDNEY, N S W

CANADA
CLARKE, IRWIN & Co. Ltd.,
480-486 University Avenue,
TORONTO.

INDIA
LONGMANS, GREEN & Co. Ltd.,
BOMBAY : 53 Nicol Road.
CALCUTTA : Chittaranjan Avenue.
MADRAS . 36A Mount Road.

SOUTH AFRICA
H. B TIMMINS, P.O. Box 94,
CAPE TOWN

Printed in Great Britain for the University of London Press Ltd.
by Hazell, Watson and Viney Ltd., London and Aylesbury.

CONTENTS

Chapter IV. The Geometrical Picture

Chapter V. Hotelling's "Principal Components"

PART II. THE ESTIMATION OF FACTORS

Chapter VI. Estimation and the Pooling Square

Chapter VII. The Estimation of Factors by Regression

Chapter VIII. Maximizing and Minimizing the Specifics

PART III. THE INFLUENCE OF SAMPLING AND SELECTION OF THE PERSONS

Chapter IX. Sampling Error and the Theory of Two Factors

CHAPTER X. MULTIPLE-FACTOR ANALYSIS WITH FALLIBLE DATA

CHAPTER XI. THE INFLUENCE OF UNIVARIATE SELECTION ON FACTORIAL ANALYSIS

CHAPTER XII. THE INFLUENCE OF MULTIVARIATE SELECTION

PART IV. CORRELATIONS BETWEEN PERSONS

CHAPTER XIII. REVERSING THE RÔLES

Chapter XIV. The Relation between Test Factors and Person Factors

PART V. THE INTERPRETATION OF FACTORS

Chapter XV. The Definition of *g*

Chapter XVI. "Simple Structure"

Chapter XVII. Limits to the Extent of Factors

CHAPTER XVIII. THE SAMPLING OF BONDS

CHAPTER XIX. "STOP-PRESS"

LIST OF DIAGRAMS

PREFACE

The theory of factorial analysis is mathematical in nature, but this book has been written so that it can, it is hoped, be read by those who have no mathematics beyond the usual secondary school knowledge. Readers are, however, urged to repeat some at least of the arithmetical calculations for themselves.

Those who wish to understand more fully the mathematical background against which the book is written are advised to read some work on statistics, say Yule and Kendall's *Introduction to the Theory of Statistics* (Griffin) especially Chapters 6, 7, 8, and 11; and, for more advanced knowledge, 12, 13, 14, and 18. T. L. Kelley's *Statistical Method* (Macmillan, New York) has the advantage of using determinants freely. Since matrix algebra plays an increasing part in factorial theory, the really serious student should read Chapter I at least of Turnbull and Aitken's *Theory of Canonical Matrices* (Blackie), and if possible also the first halves of Turnbull's *Theory of Determinants, Matrices, and Invariants* (Blackie) and Bôcher's *Introduction to Higher Algebra* (Macmillan, New York).

Those who carry out actual factorial analyses will find it almost essential to have tabular and mechanical assistance with the arithmetic. A desk slide-rule is helpful, especially in checking, and Barlow's *Tables of Squares*, etc., and Crelle's *Calculating Tables* very desirable. But any psychological laboratory doing much factorial work should have a calculating machine, one on which, for example, a tetrad-difference can be calculated without the need of noting any intermediate steps.

Even professional mathematicians will, it is hoped, read not merely the appendix, but the text. An explanation directed to the non-professional layman, and couched mainly in geometrical terms, may have suggestions for the expert also, and by being more general may counteract the

expert's alleged tendency to see one aspect of the problem too exclusively.

References to scientific articles are given thus: (Burt, 1937*b*, 84), i.e. page 84 of the second article by Burt in 1937 given in the list at the end of this book. The two important books by Spearman and by Thurstone are, however, referred to throughout by the short titles *Abilities* and *Vectors* respectively.

This book has been written during a year devoted to study and research. My sincere thanks are due to the University of Edinburgh and the Scottish National Committee for the Training of Teachers for the leave of absence, on terms than which nothing could be more generous, and to my Depute Dr. Archibald Milne and the members of staff who carried out my duties. I have also to thank warmly the Carnegie Corporation of New York for a very substantial grant, made through the Carnegie Foundation for the Advancement of Teaching and the International Institute of Teachers College, Columbia University, which has made this and other studies of the year possible under most favourable conditions.*

I am indebted to Mr. W. G. Emmett, who read a part of the MS., and to Dr. W. Ledermann, who read it all, for a number of suggestions and corrections. Among so much arithmetical work it is to be feared that some errors may still remain, for which I apologize in advance.

It is probable that the subject-matter of this book may seem to teachers and administrators to be far removed from contact with the actual work of schools. I would like therefore to explain that the incentive to the study of factorial analysis comes in my case very largely from the practical desire to improve the selection of children for higher education. When I was thirteen years of age and finishing an elementary school education, I won a " scholarship " to a secondary school in the neighbouring town, one of the early precursors of the present-day " free places "

* Carnegie Corporation is not, however, the author, owner, publisher, or proprietor of this publication, and is not to be understood as approving by virtue of its grants any of the statements made or views expressed therein.

in England. I have ever since then been greatly impressed by the influence that event has had on my life, and have spent a great deal of time in endeavouring to improve the methods of selecting pupils at that stage and in lessening the part played by chance. I take part as examiner or consultant, or as the author of tests (in co-operation with my assistants), in the conduct of many such examinations in Great Britain involving about 160,000 children every year—all the fees and royalties from which, I may perhaps be permitted to add, are devoted to financing research into the improvement of such examinations or other methods of selection. It was inevitable that I should be led to inquire into the use of intelligence tests for this purpose, and inevitable in due course that the possibilities of factorial analysis should also come under consideration. It seemed to me that before any practical use could be made of factorial analysis a very thoroughgoing examination of its mathematical foundations was necessary. The present book is my attempt at this, and as I wish to reach as many workers in this field as possible I have kept the *formulæ* of mathematics out of it as far as I could. It may seem remote from school problems. But much mathematical study and many calculations have to precede every improvement in engineering, and it will not be otherwise in the future with the social as well as with the physical sciences.

GODFREY H. THOMSON.

MORAY HOUSE,
UNIVERSITY OF EDINBURGH,
November 1938.

PART I

THE ANALYSIS OF TESTS

To simplify and clarify the exposition, errors due to sampling the population of persons are in Parts I and II assumed to be non-existent.

CHAPTER I

THE THEORY OF TWO FACTORS

1. *Factor tests.*—The object of this book is to give some account of the "factorial analysis" of ability, as it is called. In actual practice at the present day this science is endeavouring (with what hope of success is a matter of keen controversy) to arrive at an analysis of mind based on the mathematical treatment of experimental data obtained from tests of intelligence and of other qualities, and to improve vocational and scholastic advice and prediction by making use of this analysis in individual cases. It is a development of the "testing" movement—the movement in which experimenters endeavour to devise tests of intelligence and other qualities in the hope of sorting mankind, and especially children, into different categories for various practical purposes; educational (as in directing children into the school courses for which they are best suited); administrative (as in deciding that some persons are so weak-minded as to need lifelong institutional care), or vocational, etc.

There are many psychologists who would deny that from the scores in such tests, or indeed from any analysis, we can (ever) return to a full picture of the individual; and without entering into any discussion of the fundamental controversy which this denial reveals, everyone who has had anything to do with tests will readily agree that this is certainly so at present in practice. But the tester may be allowed to try to make his modest diagram of the individual better, more useful, and if possible simpler.

Now, the broadest fact about the results of "tests" of all sorts, when a large number of them is given to a large number of people, is that every individual and every test is different from every other, and yet that there are certain rather vague similarities which run through groups of people or groups of tests, not very well marked off from

one another but merging imperceptibly into neighbouring groups at their margins. To describe an individual accurately and completely one would have to administer to him all the thousand and one tests which have been or may be devised, and record his score in each, an impossible plan to carry out, and an unwieldy record to use even if obtained. Both practical necessity and the desire for theoretical simplification lead one to seek for a few tests which will describe the individual with sufficient accuracy, and possibly with complete accuracy if the right tests can be found. If, as has been said, there is some tendency for the tests to fall into groups, perhaps one test from each group may suffice. Such a set of tests might then be said to measure the "factors" of the mind.

2. *Fictitious factors.*—Actually the progress of the "factorial" movement has been rather different, and the factors are not real but as it were fictitious tests which represent certain aspects of the whole mind. But conceivably it might have taken the more concrete form. In that case the "factor tests" finally decided upon (by whom, the reader will ask, and when "finally"?) would be a set of standards which, like any other standards, would have to be kept inviolate, and unchanged except at rare intervals and for good reasons. Some tendency towards this there has been. The Binet scale of tests is almost an international standard, and there is a general agreement that it must not be changed except by certain people upon whose shoulders Binet's mantle has fallen, and only seldom and as little as possible even by them. But the Binet scale is a very complex entity, and rather represents many groups of tests than any one test. By "factor tests" one would more naturally mean tests of a "pure" nature, differing widely from one another so as to cover the whole personality adequately. And since actual tests always are more or less mixed, it is understandable why "factors" have come to be fictitious, not real, tests, to be each approximated to by various combinations of real tests so weighted that their unwanted aspects tend to cancel out, and their desired aspects to reinforce one another, the team approximating to a measure of the pure "factor."

But how, the reader will ask, do we know a "pure" factor, how are we to tell when the actual tests approximate to it? To give a preliminary answer to that question we must go back to the pioneer work of Professor Charles Spearman in the early years of this century (Spearman, 1904). The main idea which still, rightly or wrongly, dominates factorial analysis was enunciated then by him, and practically all that has been done since has been either inspired or provoked by his writings. His discovery was that the "coefficients of correlation" between tests tend to fall into "hierarchical order," and he saw that this could be explained by his famous "Theory of Two Factors." These technical terms we must now explain.

3. *Hierarchical order.*—A coefficient of correlation is a number which indicates the degree of resemblance between two sets of marks or scores. If a schoolmaster, for example, gives two examination papers to his class, say (1) in arithmetic and (2) in grammar, he will have two marks for every boy in the class. If the two sets of marks are identical the correlation is perfect, and the correlation coefficient, denoted by the symbol r_{12}, is said to be $+1$. If by some curious chance the one list of marks is exactly like the other one upside down (the best boy at arithmetic being worst at grammar, and so on), the correlation is still perfect, but negative, and $r_{12} = -1$. If there is absolutely no resemblance between the two lists, $r_{12} = 0$. If there is a strong resemblance, but falling short of identity, r_{12} may equal ·9; and so on. There is a method due to Karl Pearson of calculating such coefficients, given the list of marks.* "Tests" can obviously be correlated just like

* His "product-moment formula" is—

$$r_{12} = \frac{\text{sum }(x_1x_2)}{\sqrt{\{\text{sum }(x_1^2) \times \text{sum }(x_2^2)\}}}$$

where x_1 and x_2 are the scores in the two tests, measured from the average (so that approximately half the scores are negative), and the sums are over the persons to whom the scores apply. The quantity—

$$\sigma_1^2 = \frac{\text{sum }(x_1^2)}{\text{number of persons}}$$

is called the *variance* of Test 1, and σ_1 its *standard deviation*. If the scores in each test are not only measured from their average, but

examinations, and a convenient form in which to write down the intercorrelations of a number of tests is in a square chequer board with the names of the tests (say *a*, *b*, *c* . . .) written along the two margins, thus :

	a	*b*	*c*	*d*	*e*	*f*
a	.	·48	·24	·54	·42	·30
b	·48	.	·32	·72	·56	·40
c	·24	·32	.	·36	·28	·20
d	·54	·72	·36	.	·63	·45
e	·42	·56	·28	·63	.	·35
f	·30	·40	·20	·45	·35	.
Totals	1·98	2·48	1·40	2·70	2 24	1·70

It was early found that such correlations tend to be positive, and it is of some interest to see which of a number of tests correlates most with the others. This can be found by adding up the columns of the chequer board, when we see in the above example that the column referring to Test *d* has the highest total (2·70). The tests can then be rearranged and numbered in the order of these totals, thus :

		1	2	3	4	5	6
		d	*b*	*e*	*a*	*f*	*c*
1	*d*	.	·72	·63	·54	·45	·36
2	*b*	·72	.	·56	·48	·40	·32
3	*e*	·63	·56	.	·42	·35	·28
4	*a*	·54	·48	·42	.	·30	·24
5	*f*	·45	·40	·35	·30	.	·20
6	*c*	·36	·32	·28	·24	·20	.

After the tests have been thus arranged, the tendency which Professor Spearman was the first to notice, and which

are then divided through by their standard deviation, they are said to be *standardized*, and we represent them by z_1 and z_2. About two-thirds of them, then, lie between plus and minus one. With such scores Pearson's formula becomes—

$$r_{12} = \frac{\text{sum of the products } z_1 z_2}{\text{number of persons}}$$

In theoretical work, an even larger unit than the standard deviation is used, namely $\sigma\sqrt{p}$, where p is the number of persons. When these units are employed, the scores are said to be *normalized*. With these, the sum of the squares is unity, and the sum of the products is the correlation coefficient.

he called "hierarchical order," is more easily seen. It is the tendency for the coefficients in any two columns to have a constant ratio throughout the column. Thus in our example, if we fix our attention on Columns *a* and *f*, say, they run (omitting the coefficients which have no partners) thus:

·54	·45
·48	·40
·42	·35
.	.
.	.
·24	·20

and every number on the right is five-sixths of its partner on the left.

Our example is a fictitious one, and the tendency to hierarchical order in it has been made perfect in order to emphasize the point. It must not be supposed that the tendency is as clear in actual experimental data. Indeed, at the time there were some who denied altogether the existence of any such tendency in actual data. Those who did so were, however, mistaken, although the tendency is not as strong as Professor Spearman would seem originally to have thought (Spearman and Hart, 1912). The following is a small portion of an actual table of correlation coefficients* from those days (Brown, 1910, 309). (Complete tables must, of course, include many more tests: in recent work as many as 57 in one table.)

	1	2	3	4	5	6
1	.	·78	·45	·27	·59	·30
2	·78	.	·48	·28	·51	·24
3	·45	·48	.	52	·40	·38
4	·27	·28	·52	.	·41	·38
5	·59	·51	·40	·41	.	·18
6	·30	·24	·38	·38	·18	.

* In this, as in other instances where data for small examples are taken from experimental papers, neither criticism nor comment is in any way intended. Illustrations are restricted to few tests for economy of space and clearness of exposition, but in the experiments from which the data are taken many more tests are employed, and the purpose may be quite different from that of this book.

4. *G saturations.*—This tendency to "hierarchical order" was explained by Professor Spearman by the hypothesis that all the correlations were due to one "factor" only, present in every test, but present in largest amount in the test at the head of the hierarchy. This factor is his famous "*g*," to which he gave only this algebraic name to avoid making any suggestions as to its nature, although in some papers and in *The Abilities of Man* he has permitted himself to surmise what that nature might be. Each test had also a second factor present in it (but not to be found elsewhere, except indeed in very similar varieties of the same test), whence the name, "Theory of Two Factors"—really one general factor, and innumerable second or specific factors.

It will be proved in the Mathematical Appendix * that this arrangement would actually give rise to "hierarchical order." Meanwhile this can at least be made plausible. For if Test *d* has that column of correlations (the first in our table) with the other tests solely because it is saturated with so-and-so much *g*; and if Test *b* has less *g* in it than *d* has, it seems likely enough that *b*'s column of correlations will all be smaller in that same proportion. We can, moreover, find what these "saturations" with *g* are. For on the theory, each of our six tests contains the factor *g*, and another part which has nothing to do with causing correlation. Moreover, the higher the test is in the hierarchical ranking, the more it is "saturated" with *g*. Imagine now a fictitious test which had no specific, a test for *g* and for nothing else, whose saturation with *g* is 100 per cent., or 1·0. This fictitious test would, of course, stand at the head of the hierarchy, above our six real tests, and its row of correlations with each of those tests (their "saturations") would each be larger than any other in the same column. What values would these saturations take?

Before we answer this, let us direct our attention to the diagonal cells of the "matrix" of correlations (as it is called—a matrix is just a square or oblong set of numbers), cells which we have up to the present left blank. Since each number in our matrix represents the correlation of the two tests in whose column and row it stands, there should

* Para. 3: and see also Chapter XI, end of Section 2, page 175.

	g	1	2	3	4	5	6
g	1	r_{1g}	r_{2g}	r_{3g}	r_{4g}	r_{5g}	r_{6g}
1	r_{1g}	.	·72	·63	·54	·45	·36
2	r_{2g}	·72	.	·56	·48	·40	·32
3	r_{3g}	·63	·56	.	·42	·35	·28
4	r_{4g}	·54	·48	·42	.	·30	·24
5	r_{5g}	·45	·40	·35	·30	.	·20
6	r_{6g}	·36	·32	·28	·24	·20	.

be inserted in each diagonal cell the number *unity*, representing the correlation of a test with its own identical self. In these *self*-correlations, however, the specific factor of each test, of course, plays its part. These self-correlations of unity are the only correlations in the whole table in which specifics do play any part. These "unities," therefore, do not conform to the hierarchical rule of proportionality between the columns.

But the case is different with the fictitious test of pure g. It has no specific, and its self-correlation of unity should conform to the hierarchy. If, therefore, we call the "saturations" of the other tests r_{1g}, r_{2g}, r_{3g}, r_{4g}, r_{5g} and r_{6g}, we see that we must have, as we come down the first two columns within the matrix—

$$\frac{r_{1g}}{1} = \frac{\cdot 72}{r_{2g}} = \frac{\cdot 63}{r_{3g}} = \frac{\cdot 54}{r_{4g}} = \frac{\cdot 45}{r_{5g}} = \frac{\cdot 36}{r_{6g}}$$

a set of relations which indicate that the six "saturations" are—

$$\cdot 9 \qquad \cdot 8 \qquad \cdot 7 \qquad \cdot 6 \qquad \cdot 5 \qquad \cdot 4$$

Furthermore, each correlation in the table is the product of two of these saturations. Thus—

$$\cdot 72 = \cdot 9 \times \cdot 8$$
$$\cdot 42 = \cdot 7 \times \cdot 6$$
$$r_{34} = r_{3g} \times r_{4g}$$

The six tests can now be expressed in the form of equations—

$$z_1 = \cdot 9g + \cdot 436s_1$$
$$z_2 = \cdot 8g + \cdot 600s_2$$
$$z_3 = \cdot 7g + \cdot 714s_3$$
$$z_4 = \cdot 6g + \cdot 800s_4$$
$$z_5 = \cdot 5g + \cdot 866s_5$$
$$z_6 = \cdot 4g + \cdot 917s_6$$

Herein, each z represents the score of some person in the test indicated by the subscript, a score made up of that person's g and specific in the proportions indicated by the coefficients. The scores are supposed measured from the average of all persons, being reckoned plus if above the average and minus if below; and so too are the factors g and the specifics. And each of them, tests and factors, is "standardized," i.e. measured in such units that the sum of the squares of all the scores equals the number of persons. This is achieved by dividing the raw scores by the "standard deviation." The saturations of the specifics are such that the sum of the squares of both saturations comes in each test to unity, the whole variance of that test. Thus—

$$\cdot 436 = \sqrt{(1 - \cdot 9^2)}$$

5. *A weighted battery.*—This brief outline of the Theory of Two Factors must for the moment suffice. It is enough to enable the question to be answered which at the end of our Section 2 led to the digression. "How," the reader asked, "do we know a pure factor, how are we to tell when the actual tests approximate to it?" In the Two-factor Theory the important pure factor was g itself, and a test approximated to it the more, the higher it stood in the hierarchy. Its accuracy of measurement of g was indicated by its "saturation." And a battery of hierarchical tests could be weighted so as to have a combined saturation higher than that of any one member, each test for this purpose being weighted (as will be shown in Chapter VII) by a number proportional to $\dfrac{r_{ig}}{1 - r_{ig}^2}$, where r_{ig} is the g saturation of Test i (*Abilities*, p. xix). Although g remained a fiction, yet a complex test, made up of a weighted battery of tests which were hierarchical, could approach nearer and nearer to measuring it exactly, as more tests were added to the hierarchy. Each test added would have to conform to the rule of proportionality in its correlations with the pre-existing battery. If it did not do so it would have to be rejected. The battery at any stage would form a kind of definition of g, which it ap-

proached although never reached. And a man's weighted score in such a battery would be an estimate of *his* amount of *g*, his general intelligence. The factorial description of a man was at this period confined to one factor, since the specific factors were useless as description of any man. For one thing, they were innumerable; and for another, being specific, they were only able to indicate how the man would perform in the very tests in which, as a matter of fact, we knew exactly how he *had* performed.

6. *Oval diagrams.*—It is convenient at this point to introduce a diagrammatic illustration which will be useful in the less technical part of this book, although like all illustrations *it must be taken only as such, and the analogy must not be pushed too far.* If we represent the two abilities which are measured by tests by two overlapping ovals as in Figure 1, then the amount of the overlap can be made to represent the degree to which these tests are correlated. If we call the whole area of each oval the "variance" of that ability, we shall be introducing the reader to another technical term (of which a definition was given in the footnote to page 5). Here it need mean nothing more than the whole "amount" of the ability. The overlap we shall call the "*co-variance*." If the two variances are each equal to unity, then the covariance *is* the correlation coefficient. To make the diagram quantitative, we can indicate in figures the contents of each part of the variance, as in the instance shown, which gives a correlation of $\frac{6}{10}$, or ·6. If the separate parts of each variance (i.e. of each oval) do not add up to the same quantity, but to v_1 and v_2, say, then the covariance (the amount in the overlap) must be

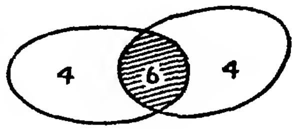

Figure 1

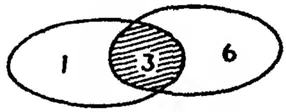

Figure 2.

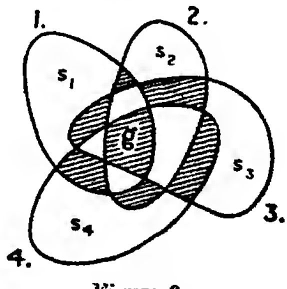

Figure 3.

divided by $\sqrt{v_1 v_2}$ in order to give the correlation. Thus, Figure 2 represents a correlation of $3 \div \sqrt{(4 \times 9)} = \cdot 5$. No attempt is made in the diagrams to make the actual areas proportional to the parts of the variance, it is the numbers written in each cell which matter.

The four abilities represented by four tests can clearly overlap in a complicated way, as in Figure 3, which shows one part of the variance (marked *g*) common to all four of the tests; four parts (left unshaded) each common to three tests; six parts (shaded) each common to two tests; and four outer parts (marked *s*) each specific to one test only. The early Theory of Two Factors adopted the hypothesis that, except for very similar varieties of the one test, none of the cells of such a diagram had any contents save those marked *g* and *s*, the general and the specific factors. The "variance" of each ability was in that theory completely accounted for by the variance due to *g*, and the variance due to *s*.

7. *Tetrad-differences.*—In Section 3 it was explained that the discovery made by Professor Spearman was that the correlation coefficients in two columns tend to be in the same ratio as we go up and down the pair of columns. That is to say, if we take the columns belonging to Tests *b* and *f*, and fix our attention on the correlations which *b* and *f* make with *d* and *e*, we have:

	b	*f*
d	·72	·45
e	·56	·35

where

$$\frac{\cdot 72}{\cdot 45} = \frac{\cdot 56}{\cdot 35}$$

This may be written—

$$\cdot 72 \times \cdot 35 - \cdot 45 \times \cdot 56 = 0$$

and in this form is called a "tetrad-difference." In symbols this one is—

$$r_{db} r_{ef} - r_{df} r_{eb} = 0$$

Spearman's discovery may therefore be put thus: "The tetrad-differences are, or tend to be, zero." It is clear that

this will be so if, as we said was the case in the Theory of Two Factors, each correlation is the product of two correlations with g. For then the above tetrad-difference becomes—

$$r_{dg}r_{bg}r_{eg}r_{fg} - r_{dg}r_{fg}r_{eg}r_{bg}$$

which is identically zero. The present-day test for hierarchical order in a correlation matrix is to calculate all the tetrad-differences (always avoiding the main diagonal) and see if they are sufficiently small. If they are, then the correlations can be explained by a diagram of the same nature as Figure 3, by one general factor and specifics. It is, of course, not to be expected in actual experimenting that the tetrad-differences will be *exactly* zero; no experiment on human material can be as accurate as that. What *is* required is that they shall be clustered round zero in a narrow curve, falling off steadily in frequency as zero is departed from. The number of tetrad-differences increases very rapidly as the number of tests grows, and in an actual experimental battery the tetrads are very numerous indeed. In the small portion of a real correlation table given above (page 7), with six tests, there are 45 tetrad-differences,* and in this instance they are distributed as follows (taking absolute values only and disregarding signs, which can be changed by altering the order of the tests:

From ·0000 to ·0999, 28 tetrad-differences.
From ·1000 to ·1999, 13 tetrad-differences.
From ·2000 to ·2796, 4 tetrad-differences.

This distribution of tetrads can be represented by a "histogram" like that shown in Figure 4, which explains itself. It is clear that some criterion is required by which we can know whether the distribution of tetrad-differences, after they have been calculated, is narrow enough to justify us in assuming the Theory of Two Factors. This criterion is explained in Part III of this book. One form of it consists in drawing a distribution curve to which, on grounds of sampling, the tetrad-differences may be expected to conform. Any tetrad-differences which seem to be too large

* Not all independent.

to be accounted for by the Theory of Two Factors are then examined, to see whether the tests giving them have any special points of resemblance, in content, method, or otherwise, which may explain why they disturb the hierarchy.

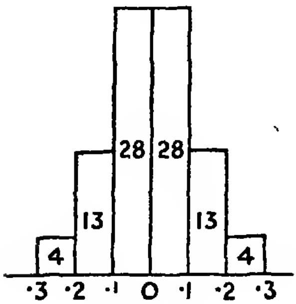

Figure 4.

8. *Group factors.*—As time went on it became clear that the tendency to zero tetrad-differences, though strong, was not universal enough to permit an explanation of *all* correlations between tests in terms of *g* and specifics, with a few slight "disturbers" in the form of slightly overlapping specifics. It became necessary to call in *group factors*, which run through many though not through all tests, to explain the deviations from strict hierarchical order. The Spearman school of experimenters, however, tend always to explain as much as possible by one central factor, and to use group factors only when necessitated. They take the point of view that a group factor must as it were establish its right to existence, that the onus of proof is on him who asserts a group factor. As a tiny artificial illustration, a matrix of correlation coefficients:

	1	2	3	4
1	.	·5	·5	·5
2	·5	.	·8	·5
3	·5	·8	.	·5
4	·5	·5	·5	.

would be examined, and its three tetrad-differences found to be:

zero
·15
·15

Inspection shows that the correlation r_{23} is the cause of the discrepancies from zero, and the experimenter trained in

the Two-factor school would therefore explain these correlations by a central factor running through them all, plus a special link joining Tests 2 and 3, as in Figure 5.

There are innumerable other possible ways of explaining these same correlations. For example, the linkages between the tests might be as in Figure 6, which gives exactly the same correlations. This lack of uniqueness is something which must always be borne in mind in studying factorial analysis. There are always, as here, innumerable possible analyses, and the final decision between them has to be made on some other grounds. The decision may be psychological, as when for example in the above case an experimenter chooses one of the possible diagrams because it best agrees with his psychological ideas about the tests. Or the decision may be made on the ground that we should be parsimonious in our invention of "factors," and that where one general and one group factor will serve we should not invent five group factors as required by Figure 6. Both diagrams, however, fit the correlational facts exactly, and so also would hundreds of other diagrams which might be made. As has been said, the two-factor tendency is to take the diagram with the largest general factor (and the largest specifics also) and with as few group factors as possible.

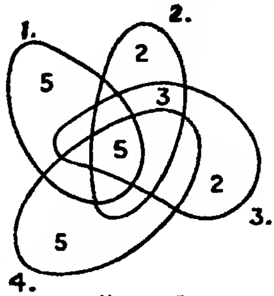

Figure 5.

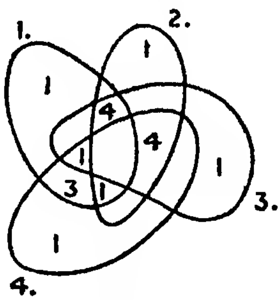

Figure 6.

9. *The verbal factor.*—In this way the Theory of Two Factors has gradually extended the "two" to include, in addition to *g* and specifics, a number of other group factors, still, however, comparatively few. These group factors bear such names as the *verbal* factor *v*, a *mechanical* factor *m*, an *arithmetic* factor, *perseveration*, etc. The charac-

teristic method of the Two-factor school can be well seen, without any technical difficulties unduly obscuring the situation, in the search for a verbal factor. The idea that, in addition to a man's g (which is generally thought of as something innate) there may be an acquired factor of verbal facility which enables him to do well in certain tests, is a not unnatural one. A battery of tests can be assembled, of which half do, and half do not, employ words in their construction or solution. The correlation matrix will then have four quadrants, the quadrant V containing the correlations of the verbal tests among themselves, the

$$\begin{array}{|cc|} V & C \\ \\ C & P \end{array}$$

quadrant P the correlations of the non-verbal or, say, pictorial tests, and the quadrants C containing the cross-correlations of the one kind of test with the other. If the whole table is sufficiently "hierarchical," there is no evidence for a group factor v or a group factor p. If either of these factors exists, there will be differences to be noticed between the six kinds of tetrad which can be chosen, namely:

$$\begin{array}{c|cc|} & p & p \\ v & \cdot & \cdot \\ & \multicolumn{2}{c|}{(1)} \\ v & \cdot & \cdot \end{array} \qquad \begin{array}{c|cc|} & v & v \\ v & x & x \\ & \multicolumn{2}{c|}{(2)} \\ v & x & x \end{array} \qquad \begin{array}{c|cc|} & p & p \\ p & x & x \\ & \multicolumn{2}{c|}{(3)} \\ p & x & x \end{array}$$

$$\begin{array}{c|cc|} & v & p \\ v & x & \cdot \\ & \multicolumn{2}{c|}{(4)} \\ v & x & \cdot \end{array} \qquad \begin{array}{c|cc|} & v & p \\ p & \cdot & x \\ & \multicolumn{2}{c|}{(5)} \\ p & \cdot & x \end{array} \qquad \begin{array}{c|cc|} & v & p \\ v & x & \cdot \\ & \multicolumn{2}{c|}{(6)} \\ p & \cdot & x \end{array}$$

A tetrad like 1, with two verbal tests along one margin and two pictorial tests along the other, will be found in

quadrant *C*. Neither a factor common to the verbal tests only, nor one common to the pictorial tests only, will add anything to any of the four correlations in such a tetrad-difference, which may be expected, therefore, to tend to be zero. If the tetrads in *C* seem to do so, the other tetrads can be examined. Tetrad 2 is taken wholly from the *V* quadrant. In it the verbal factor, if any is present, will reinforce all the four correlations, and should not therefore disturb very much the tendency to a zero tetrad-difference. (Reinforced correlations are marked by x in the diagrams.) The same is true of Tetrad 3 taken wholly from the *P* quadrant. Tetrads 4 and 5 have each two of their correlations reinforced, by the v factor in 4 and by the p factor in 5, but in each case in such a way as not to change very much the tetrad-difference. It is when we come to tetrads like 6, which have one correlation in each of the four quadrants, that the presence of either or both factors should show itself strongly: for the two reinforced correlations here occur on a diagonal, and inflate only the one member of the tetrad-difference—

$$r_{vv}r_{pp} - r_{vp}r_{pv}$$

If, then, a verbal factor, and also a pictorial factor, are present, the tendency for the tetrad-differences to vanish should become less and less strong as we consider tetrads of the kinds 1, 2 and 3, 4 and 5, and especially 6, where the tetrad-differences should leap up. If only the verbal factor is present, tetrad-differences of the kind 3 should vanish rather more than those of the kind 2. But it will not be easy to distinguish between either suspected factor, and both. Tetrads like 6, however, should give conclusive evidence of the presence of one or the other, if not both. Methods like this were employed by Miss Davey (Davey, 1926), who found a group factor, but not one running through *all* the verbal tests, and by Dr. Stephenson (Stephenson, 1931), whose results indicated the presence of a verbal factor.*

10. *Group-factor saturations.*—Just as the g saturations

* T. L. Kelley had already found by other methods strong evidence of a verbal factor (Kelley, 1928, 104, 121 *et passim*).

of tests can be calculated, so also can the saturation of a test with any group factor it may contain. The general method of the Two-factor school is first to work with batteries of tests which give no unduly large tetrad-differences, and which also appear to satisfy one's general impression that they test intelligence. From such a battery, of which the best example is that of Brown and Stephenson (B. and S., 1933), the g saturations can be calculated.* Each test has, however, also its specific, which, *so long as it is in the hierarchical battery*, is unique to it and shared with no other member of the battery. A test may now be associated with some other battery of different tests, and with some of these it may share a part of its former specific, as a group factor which will increase its correlation beyond that caused by g. The excess correlation enables the saturation of the test with this group factor to be found—the details are too technical for this chapter—and the specific saturation correspondingly reduced. Finally, the tester may be able to give the composition of a test as, let us say (to invent an example)—

$$\cdot 71g + \cdot 40v + \cdot 34n + \cdot 47s$$

where g is Spearman's g, v is Stephenson's verbal factor, n is a number factor, and s is the remaining specific of the test. The coefficients are the "saturations" of the test with each of these; that is, the correlations believed to exist between the test and these fictitious tests called factors. The squares of these saturations represent the fractions of the test-variance contributed by each factor, and these squares sum to unity, thus:

	Saturation Squared
g . . .	·5041
v . . .	·1600
n . . .	·1156
s . . .	·2209
	1·0006

* For the sake of clarity the text here rather oversimplifies the situation. The battery of Brown and Stephenson contains in fact a rather large group factor as well as g and specifics.

11. *The bifactor method.*—Holzinger's Bifactor Method (Holzinger, 1935, 1937*a*) may be looked upon as another natural extension of the simple Two-factor plan of analysis. It endeavours to analyse a battery of tests into one general factor running through all of them, and a number of *mutually exclusive* group factors each of which runs through a group only. A diagram of such an analysis looks like a "hollow staircase," thus:

Test	*g*	*h*	*k*	*l*
1	×	×		
2	×	×		
3	×		×	
4	×		×	
5	×			×
6	×			×

Here factor *g* runs through all, as is indicated by the column of crosses. Factors *h*, *k*, and *l* run through small and mutually exclusive groups of tests each. The saturations with *g* can be calculated from sub-batteries of tests which form perfect hierarchies, by selecting only one test from each group (in every possible way). After these are known, the correlation due to *g* can be removed, and then the saturations due to each group factor found, for which purpose, however, more tests than two would ordinarily be required in each group—our diagram is restricted to two only for simplicity and economy of space.

12. *Vocational guidance.*—It will clearly be an aim of the experimenter along all these lines to obtain if possible single real tests, or failing that weighted batteries of tests, which approximate as closely as possible to the factors he has found, or postulated; and with these to estimate the amount of each factor possessed by any man, and also (by giving such tests to tried workmen or school pupils) to estimate the amount of each factor required by different "occupations" (including higher education) with a view to vocational and educational selection and guidance.

CHAPTER II

MULTIPLE-FACTOR ANALYSIS

1. *Need of group factors.*—The two-factor method of analysis, described in the last chapter, began with the idea that a matrix of correlations would ordinarily show perfect hierarchical order if care was taken to avoid tests which were "unduly similar," i.e. very similar indeed to one another. If such were found coexisting in the team of tests, the team had to be "purified" by the rejection of one or other of the two. Later it became clear that this process involves the experimenter in great difficulty, for it subjects him to the temptation to discover "undue similarity" between tests *after* he has found that their correlation breaks the hierarchy. Moreover, whole groups of tests were found to fail to conform; and so group factors were admitted, though always, by the experimenter trained in that school, with reluctance and in as small a number as possible. It had, however, become quite clear that the *Theory* of Two Factors in its original form had been superseded by a theory of many factors, although the *method* of two factors remained as an analytical device for indicating their presence and for isolating them in comparative purity.

Under these circumstances it is not surprising that some workers turned their attention to the possibility of a method of multiple-factor analysis, by which any matrix of test correlations could be analysed direct into its factors (Garnett, 1919*a* and *b*). It was Professor Thurstone of Chicago who saw that one solution to this problem could be reached by a generalization of Spearman's idea of zero tetrad-differences.

2. *Rank of a matrix and number of factors.*—We saw that when all the tetrad-differences are zero, the correlations can all be explained by *one* general factor, a tetrad being

formed of the intercorrelations of two tests with two other tests, thus:

	3	4
1	r_{13}	r_{14}
2	r_{23}	r_{24}

and the tetrad-difference being—

$$r_{13}r_{24} - r_{23}r_{14}$$

Thurstone's idea, though rather differently expressed by him (*Vectors*, Chapter II), can be based on a second, third, fourth . . . calculation of certain tetrad-differences of tetrad-differences.

To explain this, let us consider the correlation coefficients which three tests make with three others:

	4	5	6
1	r_{14}	r_{15}	r_{16}
2	r_{24}	r_{25}	r_{26}
3	r_{34}	r_{35}	r_{36}

This arrangement of nine correlation coefficients might have been called a "nonad," by analogy with the tetrad. Actually, by mathematicians, it is called a "minor determinant of order three" or more briefly a three-rowed minor; a tetrad is in this nomenclature a "minor of order two."

We can now, on the above three-rowed determinant, perform the following calculation. Choose the top left coefficient as "pivot," and calculate the four tetrad-differences of which it forms part, namely:

$$(r_{14}r_{25} - r_{24}r_{15}) \qquad (r_{14}r_{26} - r_{24}r_{16})$$
$$(r_{14}r_{35} - r_{34}r_{15}) \qquad (r_{14}r_{36} - r_{34}r_{16})$$

These four tetrad-differences now themselves form a tetrad which can be evaluated. If it is zero, we say that the three-rowed determinant with which we started "vanishes."

Exactly the same repeated process can be carried on with larger minor determinants. For example, the minor of order four here shown vanishes:

	(·26)	·32	·88	·84
	·42	·36	·62	·72
	·44	·62	·66	·46
	·45	·58	·63	·60
for its pivotal t.d.'s are		(− ·0408)	·0016	·0444
		·0204	·0044	− ·0300
		·0068	− ·0072	·0030
and then			(− ·00021216)	·00031824
			·00028288	− ·00042432
and finally			zero	

This process of continually calculating tetrads is called "pivotal condensation." The reader should be given a word of warning here, that the end result of this form of calculation, if not zero, has to be divided by the product of all the pivots except the last, to give the value of the determinant we began with. A routine method (Aitken, 1937*a*) of carrying out pivotal condensation, including division by the pivot at each step, is described in Chapter VI, page 89 ff.*

We can in this way examine the minors of order two, three, four (and so on) of a correlation matrix, always avoiding those diagonal cells which correspond to the correlation of a test with itself. We may come to a point at which *all* the minors *of that order* vanish. Suppose these minors which all vanish are the minors of order five. We then say that the " rank " of the correlation matrix is four (with the exception of the diagonal cells). There then exists the possibility that the " rank " of the *whole* correlation matrix can be reduced to four by inserting suitable quantities in the diagonal cells (see next section). The " rank " of a matrix is the order of its largest† non-vanish-

* If the process gives, at an earlier stage than the end, a matrix entirely composed of zeros, the rank of the original determinant is correspondingly less, being equal to the number of condensations needed to give zeros.

† " Largest " refers to the number of rows, not to the numerical value.

ing minor. Thurstone's discovery was that the tests could be analysed into as many common factors as the above reduced rank of their correlation matrix—the rank, that is to say, apart from the diagonal cells—plus a specific in each test. He also invented a method of performing the analysis.

3. *Thurstone's method used on a hierarchy.*—Thurstone's rule about the rank includes Spearman's hierarchy as a special case, for in a hierarchy the tetrads—that is, the minors of order *two*—vanish. The rank is therefore *one*, and a hierarchical set of tests can be analysed into *one* common factor plus a specific in each. A simple way of introducing the reader to Thurstone's hypothesis and also to his "centroid" method * of finding a set of factor saturations will be to use it first of all on the perfect Spearman hierarchy which we cited as an artificial example in our first chapter.

Tests	1	2	3	4	5	6
1	.	·72	·63	·54	·45	·36
2	·72	.	·56	·48	·40	·32
3	·63	·56	.	·42	·35	·28
4	·54	·48	·42	.	·30	·24
5	·45	·40	·35	·30	.	·20
6	·36	·32	·28	·24	·20	.

The first step in Thurstone's method, after the rank has been found, is to place in the blank diagonal cells numbers which will cause these cells also to partake of the same rank as the rest of the matrix, numbers which, for a reason which will become clear later, are called "communalities." In our present Spearman example that rank is *one*, i.e. the tetrads vanish. The communalities, therefore, must be such numbers as will make also those tetrads vanish which include a diagonal cell: this enables them to be calculated. Let us, for example, fix our attention on the communality of the first test, which we will designate h_1^2 (the reason for the "square" will become apparent later). Then the tetrad formed by Tests 1 and 2 with Tests 1 and 3 is:

* We shall see why it is called the "centroid" method in Section 9 of Chapter VI, after we have learned to use a "pooling square."

	1	3
1	h_1^2	·63
2	·72	·56

and the tetrad-difference has to vanish. Therefore—

$$\cdot 56 h_1^2 - \cdot 72 \times \cdot 63 = 0$$
$$\therefore h_1^2 = \cdot 81$$

Similarly all the communalities can be calculated, and found to be—

·81 ·64 ·49 ·36 ·25 ·16

(The observant reader will notice that they are the squares of the "saturations" of our first chapter; but let us continue with Thurstone's method as though we had not noticed this.)

Thurstone's method of finding the saturations of each test with the first common factor is then to insert the communalities in the diagonal cells and add up the columns * of the matrix, thus:

Original Correlation Matrix

(·81)	·72	·63	·54	·45	·36	
·72	(·64)	·56	·48	·40	·32	
·63	·56	(·49)	·42	·35	·28	
·54	·48	·42	(·36)	·30	·24	
·45	·40	·35	·30	(·25)	·20	
·36	·32	·28	·24	·20	(·16)	
3·51	3·12	2·73	2·34	1·95	1·56	15·21

The column totals are then themselves added together (15·21) and the square root taken (3·90). The "satura-

* This, the "centroid" method of finding a set of loadings, is not in any way bound up with Thurstone's theorem about the rank and the number of common factors. It can be used, for example, with unity in each diagonal cell, in which case it will give as many factors as there are tests and saturations somewhat resembling those given by Hotelling's process described in Chapter V: and vice versa Hotelling's process could be used on the matrix with communalities inserted.

tions " of the first (and here the only) common factor are then the columnar totals divided by this square root, namely—

$$\frac{3{\cdot}51}{3{\cdot}90} \quad \frac{3{\cdot}12}{3{\cdot}90} \quad \frac{2{\cdot}73}{3{\cdot}90} \quad \frac{2{\cdot}34}{3{\cdot}90} \quad \frac{1{\cdot}95}{3{\cdot}90} \quad \frac{1{\cdot}56}{3{\cdot}90}$$

or $\quad {\cdot}9 \quad {\cdot}8 \quad {\cdot}7 \quad {\cdot}6 \quad {\cdot}5 \quad {\cdot}4$

as in the present instance we already know them to be. (Very often in multiple-factor analysis the " saturation " of a test with a factor is called the " loading," and this is a convenient place to introduce the new term.)

As applied to the hierarchical case, this method of finding the saturations or loadings had been devised and employed many years previously by Cyril Burt, though it is not quite clear how he would have filled in the blank diagonal cells (Burt, 1917, 53, footnote). It should also be explained to the reader that in actual practice Thurstone and his followers do not calculate the minor determinants to find the rank and the communality, for that would be too laborious. Instead they adopt the approximation of inserting in each diagonal cell the largest correlation coefficient of the column (see Chapter X).

4. *The second stage of the " centroid " method.*—If there is more than one common factor, the process goes on to another stage. Even with our example we can show the beginning of this second stage, which consists in forming that matrix of correlations which the first factor alone would produce. This is done by writing the loadings along the two sides of a chequer board and filling every cell of the chequer board with the product of the loading of that row with the loading of that column, thus :

First-factor Matrix

	·9	·8	·7	·6	·5	·4
·9	·81	·72	·63	·54	·45	·36
·8	·72	·64	·56	·48	·40	·32
·7	·63	·56	·49	·42	·35	·28
·6	·54	·48	·42	·36	·30	·24
·5	·45	·40	·35	·30	·25	·20
·4	·36	·32	·28	·24	·20	·16

This is the "first-factor matrix," which gives the parts of the correlations due to the first factor. This matrix has now to be subtracted from the original matrix to find the residues which must be explained by further common factors.

In our present example the first-factor matrix is identical with the original matrix and *the residues are all zero.* Only the one common factor is therefore required. (Of course, the reader will understand that in a real experimental matrix the residues can never be expected to be *exactly* zero: one is content when they are near enough to zero to be due to chance experimental error.) Had the rank of our original matrix of correlations been, however, higher than *one*, there would have been a matrix of residues.

Let us now make an artificial example with a larger number of common factors, say *three*, which we can afterwards use to illustrate the further stages of Thurstone's method. We can do this in an illuminating manner by the aid of the oval diagrams described in Chapter I.

5. *A three-factor example.*—In Figure 7, a diagram of the overlapping variances of four tests, let us insert three common factors and specifics to complete the variance of each test to 10 (to make our arithmetical work easy). No factor here is common to all the four tests. The factor with a variance of 4 runs through Tests 1, 2, and 3. That with a variance 3 runs through Tests 2, 3, and 4. That with a variance 2 runs through Tests 1 and 4. The other factors are specifics. The four test variances being each 10, the correlation coefficients are written down from the overlaps by inspection as:

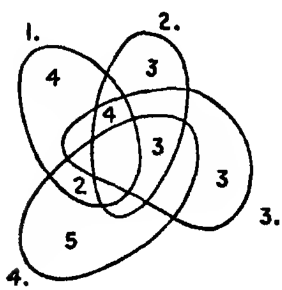

Figure 7.

	1	2	3	4
1	(·6)	·4	·4	·2
2	·4	(·7)	·7	·3
3	·4	·7	(·7)	·3
4	·2	·3	·3	(·5)

Moreover, we can put into our matrix the communalities corresponding to our diagram. Each communality is, in fact, that fraction of the variance of a test which is not specific. Thus ·6 of the variance of Test 1 is "communal," ·4 being specific or "selfish." In this way we have the matrix above, with communalities inserted. We can now pretend that it is an experimental matrix, ready for the application of Thurstone's method, as follows :

	(·6)	·4	·4	·2	
	·4	(·7)	·7	·3	Original
	·4	·7	(·7)	·3	experimental
	·2	·3	·3	(·5)	matrix.
	1·6	2·1	2·1	1·3	= 7·1 = 2·6646²
1st Loadings	·6005	·7881	·7881	·4879	= 2·6646 *

·6005	(·3604)	·4733	·4733	·2930	
·7881	·4733	(·6211)	·6211	·3845	First-factor
·7881	·4733	·6211	(·6211)	·3845	matrix.
·4879	·2930	·3845	·3845	(·2380)	

Here it is seen that the loadings of the first factor, when cross-multiplied in a chequer board, give a first factor matrix which is *not* identical with the original experimental matrix, unlike the case of the former, hierarchical, matrix. Here (as we who made the matrix know) one factor will not suffice. We subtract the first-factor matrix from the original experimental matrix to see how much of the correlations still has to be explained, and how much of the "communalities" or communal variances. The latter were—

·6 ·7 ·7 ·5

and of these amounts the first factor has explained—

·3604 ·6211 ·6211 ·2380

If we subtract the first-factor matrix, element by element, from the original experimental matrix, we get the residual matrix :

* This check should always be applied. To avoid complication it is not printed in the later tables. It applies to the loadings with their temporary signs (see below).

(·2896)	− ·0783	− ·0783	− ·0930	
− ·0783	(·0789)	·0789	− ·0845	First residual
− ·0783	·0789	(·0789)	− ·0845	matrix.
− ·0930	− ·0845	− ·0845	(·2620)	

To this matrix we are now going to apply exactly the same procedure as we applied to the original experimental matrix, in order to find the loadings of the second factor. But we meet at once with a difficulty. The columns of the residual matrix add up exactly * to zero! This always happens, and is indeed a useful check on our arithmetical work up to this point, but it seems to stop our further progress.

To get over this difficulty *we change temporarily the signs of some of the tests* in order to make a majority of the cells of each column of the matrix positive. In the present instance we could make them nearly all positive, by changing the signs of Tests 1 and 4. That is to say, we could change the signs in the first and last row, and then in the first and last column. (The four corner elements would thus have their signs first changed, and then changed back again.) The columns can be made to have mainly positive totals, however, in several different ways, as a rule, and it is desirable to have a fixed method for doing this. The practice adopted by Thurstone in *The Vectors of Mind* is to change the sign of the test with most minuses in its column and row, and so on until there is a large majority of plus signs. We shall adopt his easier rule given in *A Simplified Factor Method*, i.e. to seek out the column whose total *regardless of signs* is the largest, and then temporarily change the signs of variables so as to make all the signs in that column positive.

The sums of the above columns, regardless of sign, are—

·4792	·3156	·3156	·5240

and therefore we must change the signs of tests so as to make all the signs in Column 4 positive; that is, we must change the signs of the first three tests.† Since we change

* When enough decimals have been retained. In practice there may be a discrepancy in the last decimal place.

† Changing the sign of Test 4 would here have the same result, but for uniformity of routine we stick to the letter of the rule.

the three row signs, as well as the three column signs, this will leave a block of signs unchanged, but will make the last column and the last row all positive. We now have :

	·2396	− ·0733	− ·0733	(−)·0930	
	− ·0733	·0789	·0789	(−)·0845	First residual
	− ·0733	·0789	·0789	(−)·0845	matrix with
	(−)·0930	(−)·0845	(−)·0845	·2620	changed signs.
	·1860	·1690	·1690	·5240	$= 1{\cdot}0480$
					$= 1{\cdot}0237^2$
2nd Loadings	·1817	·1651	·1651	·5119	With temporary signs.
·1817	·0330	·0300	·0300	·0930	
·1651	·0300	·0273	·0273	·0845	Second-factor
·1651	·0300	·0273	·0273	·0845	matrix.
·5119	·0930	·0845	·0845	·2620	
	·2066	− ·1033	− ·1033	.	
	− ·1033	·0516	·0516	.	Second residual
	− ·1033	·0516	·0516	.	matrix.
	.	.	.	.	

On the matrix with these temporarily changed signs we then operate exactly as we did on the original experimental matrix,* and obtain second-factor loadings which (*with temporary signs*) are—

·1817 ·1651 ·1651 ·5119

The second-factor matrix, that is, the matrix showing how much correlation is due to the second factor, is then made on a chequer board *still using the temporary signs*, and subtracted from the previous matrix of residues (*with its temporary signs*, not with its first signs) to find the residues still remaining, to be explained by further factors. In the present instance we see that the whole variance of the fourth test entirely disappears, and also all the correlations in which that test is concerned. This test, therefore, is fully explained by the two factors already extracted. Only the first three test variances remain unexhausted, and their correlations. Again the columns of the residual

* The totals of *some* of the columns may be negative, without detriment to the process. It is the algebraic sum of the totals which is then taken, and its square root used as divisor to get the loadings.

matrix sum exactly to zero. Following our rule, the signs of Tests 2 and 3 have to be temporarily changed before the process can continue. After these changes of sign the second residual matrix is as follows, and the same operation as before is again performed on it :

	2066	(−)·1033	(−)·1033	.	Second residual matrix with signs temporarily changed.
	(−)·1033	·0516	·0516	.	
	(−)·1033	·0516	·0516	.	
	.	.	.	.	
	·4132	2065	·2065	.	·8262 − ·9090²
3rd Loadings	·1545	·2272	·2272	.	with temporary signs.

With these third-factor loadings we can now calculate the variances and correlations due to the third factor : and we find these are exactly equal to the second residual matrix. On subtracting, the third residual matrix we obtain is entirely composed of zeros. (In a practical example we should be content if it was sufficiently small.) We thus find (as our construction of the artificial tests entitled us to expect) that the matrix of correlations can be completely explained by three common factors.

After the analysis has been completed, some care is needed in returning from the temporary signs of the loadings to the correct signs. The only safe plan is to write down first of all the loadings with their temporary signs as they came out in the analysis. In our present example these happen to be all positive, though that will not always occur.

Loadings with Temporary Signs

Test	*I*	*II*	*III*
1	·6005	·1817	·1545
2	·7881	·1651	·2272
3	·7881	·1651	·2272
4	·4879	·5119	.

Now, in obtaining Loadings II the signs of Tests 1, 2, and 3 were changed. We must, therefore, in the above table reverse the signs of the loadings of these three tests in

Column II *and each later column.* Then in obtaining Loadings III the signs of Test 2 and 3 were changed; that is, in our case changed back to positive. The loadings with their proper signs are therefore as shown in the first three columns of this table:

Test	*Loadings of the Factors (Signs Replaced)*						
	I	*II*	*III*	*Specific*			
1	·6005	− ·1817	·4545	·6324	.	.	.
2	·7881	− ·1651	·2272	.	·5477	.	.
3	·7881	− ·1651	·2272	.	.	·5477	.
4	·4879	·5119	.	.	.	.	·7071

In this table each column of loadings, for the common factors after the first, adds up to zero. The loading of the specific is found from the fact that in each row the sum of the squares must be unity, being the whole variance of the test. The inner product * of each pair of rows gives the correlation between those two tests (Garnett, 1919*a*). Thus —

$$r_{12} = \cdot6005 \times \cdot7881 + \cdot1817 \times \cdot1651 - \cdot4545 \times \cdot2272 = \cdot4000$$

in agreement with the entry in the original correlation matrix. With artificial data like the present, the analysis results in loadings which give the correlations back exactly.

It will be seen that all the signs in any column of the table of loadings can be reversed without making any change in the inner products of the rows; that is, without altering the correlations. We would usually prefer, therefore, to reverse the signs of a column like our Column III, so as to make its largest member positive.

The amount which each factor contributes to the variance of the test is indicated by the square of its loading in that

* By the "inner product" of two series of numbers is meant the sum of their products in pairs. Thus the inner product of the two series:

$$\begin{matrix} a & b & c & d \\ A & B & C & D \end{matrix}$$

is

$$aA + bB + cC + dD$$

test. The sum of the squares of the three common-factor loadings gives the "communality" which we originally deduced from Figure 7 and inserted in the diagonal cells of our original correlation matrix. These facts can be better seen if we make a table of the squares of the above loadings :

Test	*Variance contributed by Each Factor*					
	I	*II*	*III*	*Communality*	*Specific Variance*	*Total*
1	·3604	·0330	·2066	·6000	·4000	1
2	·6211	·0273	·0516	·7000	·3000	1
3	·6211	·0273	·0516	·7000	·3000	1
4	·2380	·2620	.	·5000	·5000	1
Total	1·8406	·3496	·3098	2·5000	1·5000	4

6. *Comparison of the analysis with the diagram.*—The reader has probably been turning from this calculation of the factor loadings back to the four-oval diagram with which we started, in order to detect any connection ; and has been disappointed to find none. The fact is that the analysis to which the Thurstone method has led us is, except that it too has three common factors, a different analysis from that which the original diagram naturally invites. That diagram gave for the variance due to each factor the following :

Test	*Variance contributed by Each Factor*					
	I	*II*	*III*	*Communality*	*Specific Variance*	*Total*
1	·4	.	·2	·6	·4	1
2	·4	·3	.	·7	·3	1
3	·4	·3	.	·7	·3	1
4	.	·3	·2	·5	·5	1
Totals	1·2	·9	·4	2·5	1·5	4

and the factor loadings are the positive square roots of these.

Test	*Loadings of the Factors*						
	I	*II*	*III*	*Specifics*			
1	6325	.	4472	6324	.	.	.
2	·6325	·5477	.	.	·5477	.	.
3	·6325	·5477	.	.	.	·5477	.
4	.	·5477	·4472	.	.	.	·7071

The only points in common between the two analyses are that they both have the same communalities (and therefore the same specific variances) and the same number of common factors. The Thurstone analysis has two general factors (running through all four tests), while the diagram had none : and the Thurstone analysis has several negative loadings, while the diagram had none. We shall see later that Thurstone, after arriving at this first analysis, endeavours to convert it into an analysis more like that of our diagram, with no negative loadings and no completely general factors. This is one of the most difficult yet essential parts of his method.

7. *Analysis into two common factors.*—When we began our analysis of the matrix of correlations corresponding to Figure 7, we simply put the communalities suggested by that figure into the blank diagonal cells. That served to illustrate the fact that the Thurstone method of calculation will bring out as many factors as correspond to the communalities used, here three factors. But it disregarded (intentionally for the purpose of the above illustration) a cardinal point of Thurstone's theory that we must seek for the communalities which make the rank of the matrix a minimum, and therefore the number of common factors a minimum. We simply accepted the communalities suggested by the diagram. Let us now repair our omission and see if there is not a possible analysis of these tests into fewer than three common factors. There is no hope of reducing the rank to one, for the original correlations give

two of the three tetrads different from zero, and we may (in an artificial example) assume that there are no experimental or other errors. But there is nothing in the experimental correlations to make it certain that rank 2 cannot be attained. With only four tests (far too few, be it remembered, for an actual experiment) there is no minor of order three entirely composed of experimentally obtained correlations. It may then be the case that communalities can be found which reduce the rank to 2. Indeed, as we shall see presently, many sets of communalities will do so, of which one is shown here:

$$\begin{array}{cccc}
(\cdot 2\dot{6}) & \cdot 4 & \cdot 4 & \cdot 2 \\
\cdot 4 & (\cdot 7) & \cdot 7 & \cdot 3 \\
\cdot 4 & \cdot 7 & (\cdot 7) & \cdot 3 \\
\cdot 2 & \cdot 3 & \cdot 3 & (\cdot 15)
\end{array}$$

These communalities $\cdot 2\dot{6}$, $\cdot 7$, $\cdot 7$, and $\cdot 15$ make every three-rowed minor exactly zero. For example, the minor

$$\begin{array}{ccc}
(\cdot 2\dot{6}) & \cdot 4 & \cdot 2 \\
\cdot 4 & (\cdot 7) & \cdot 3 \\
\cdot 2 & \cdot 3 & (\cdot 15)
\end{array}$$

becomes by " pivotal condensation ":

$$\begin{array}{cc}
\cdot 02\dot{6} & 0 \\
0 & 0 \\
\hline
 & 0
\end{array}$$

and finally 0

It must, therefore, be possible to make a four-oval diagram, showing only two common factors, and indeed

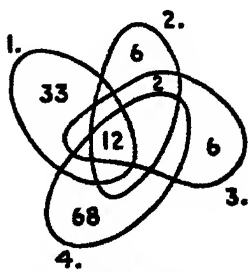

Figure 8.

more than one such diagram can be found. One is shown in Figure 8.

This gives exactly the correct correlations. For example—

$$r_{23} = \frac{12 + 2}{\sqrt{(20 \times 20)}} = \frac{14}{20} = \cdot 7$$

$$r_{34} = \frac{12}{\sqrt{(20 \times 80)}} = \frac{12}{40} = \cdot 3$$

It also gives the communalities $\cdot 2\dot{6}$, $\cdot 7$, $\cdot 7$, $\cdot 15$. For example, in Test 1, variance to the amount of 12 out of 45 is communal, and $12/45 = \cdot 2\dot{6}$.

The insertion of these communalities, therefore, in the matrix of correlations ought to give a matrix which only two applications of Thurstone's calculation should completely exhaust. The reader is advised to carry out the calculation as an exercise. He will find for the first-factor loadings—

·5000 ·8290 ·8290 ·3750

and if in the first residual matrix, following our rule, he changes temporarily the signs of Tests 2 and 3, the second-factor loadings will be—

·1291 − ·1128 − ·1128 ·0968

The second residual matrix will be found to be exactly zero in each of its sixteen cells. The variance (square of the loading) contributed by each factor to each test is then in this analysis :

Test	*Variance contributed by Each Factor*				
	I	*II*	*Communality*	*Specific Variance*	*Total*
1	·2500	·0167	·2667	·7333	1
2	·6873	·0127	·7000	·3000	1
3	·6873	·0127	·7000	·3000	1
4	·1406	·0094	·1500	·8500	1
Totals	1·7652	·0515	1·8167	2·1833	4

If we now compare these analyses, we see that the three common factors of the previous analysis " took out," as

the factorial worker says, a variance of 2·5 of the total 4, leaving 1·5 for the specifics. The present analysis leaves 2·1833 for the specifics which here form a larger part of the four tests.

8. *Rotation of the axes.*—We saw in Section 6 that the Thurstone method there led to an analysis which was different from the analysis corresponding to the diagram with which we began. That is also the case with the present analysis into two common factors—the very fact that it gives the second factor two negative loadings shows this, for the diagram (Figure 8) corresponds to positive loadings only. We said, too, in Section 6 that a difficult part of Thurstone's method was the conversion of the loadings into new and equivalent loadings which are all positive. This will form the subject of a later and more technical chapter; but a simple illustration of one method of conversion (or "rotation" as it is called, for a reason which will become clear later) can be given from our present example. It is a method which can be used only if we have reason to think that one of our tests contains only one common factor (Alexander, 1935, 144). Let us suppose in our present case that from other sources we know this fact about Test 1. The centroid analysis has given us the loadings shown in the first two columns of this table:

Test	*Unrotated Loadings*		*Communality*	*Rotated Loadings*		*Rotated Loadings*	
	I	*II*		*I**	*II**	*I***	*II***
1	·5000	·1291	·2667	·5164	.	·4781	·1952
2	·8291	− ·1128	·7000	·7746	·3162	·8367	.
3	·8291	− ·1128	·7000	·7746	·3162	·8367	.
4	·3750	·0968	·1500	·3873	.	·3586	·1464

The communalities are also shown; they are the sums of the squares of the loadings. If now we know or decide to assume that Test 1 has really only one common factor, and if we want to preserve the communalities shown, then the

loading of factor I* in Test 1 must be the square root of ·2667, namely ·5164.

The loadings of factor I* in the other three tests can now be found from the fact that they must give the correlations of those tests with Test 1, since Test 1 has no second factor to contribute. The loadings shown in column I* are found in this way: for example, ·7746 is the quotient of ·5164 divided into r_{12} (·4), and ·3873 is similarly r_{14} (·2) divided by ·5164.

The contributions of factor I* to the communalities are obtained by squaring these loadings. In Test 1, we already know that factor I* exhausts the communality, for that is how we found its loading. We discover that in Test 4, factor I* likewise exhausts the communality, for the square of ·3873 is ·1500. The other two tests, however, have each an amount of communality remaining equal to ·1000 (i.e. ·7000 − ·7746²). The square root of ·1000, therefore (·3162), must be the loading of factor II* in Tests 2 and 3. The double column of loadings ought now to give all the correlations of the original correlation matrix, and we find that it does so. Thus, e.g.—

$$r_{23} = \cdot 7746 \times \cdot 7746 + \cdot 3162 \times \cdot 3162 = \cdot 7000$$

and $$r_{24} = \cdot 7746 \times \cdot 3873 = \cdot 3000$$

Moreover, the analysis into factors I* and II* corresponds exactly to Figure 8. For example, the loading of factor II* in Test 2 in that diagram is the square root of 2/20 (·3162); and the loading of factor I* in Test 4 is the square root of 12/80 (·3873).

If, however, the experimenter had reasons for thinking that Test 2 (not Test 1) was free from the second common factor, his "rotation" of the loadings would have given a different result, shown in the table opposite in columns I** and II**. This set of loadings also gives the correct communalities and the experimental correlations, but does not correspond to Figure 8. A diagram can, however, be constructed to agree with it (Figure 9), and the reader is advised to check the agreement by calculating from the diagram the loadings of each factor, the communalities of each test, and the correlations.

We have had, in Figures 7, 8, and 9, three different analyses of the same matrix of correlations. If with Thurstone we decide that analyses must always use the minimal number of common factors, we will reject Figure 7. Between Figures 8 and 9, however, this principle makes no choice. Much of the later and more technical part of Thurstone's method is taken up with his endeavours to lay down conditions which will make the analysis unique.

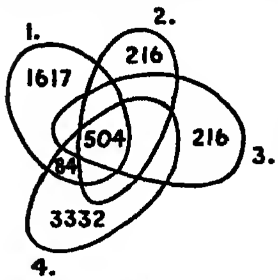

Figure 9.

9. *Unique communalities.*—The first requirement for a unique analysis is that the set of communalities which gives the lowest rank should be unique, and this is not the case with a battery of only four tests and minimal rank 2, like our example. There are many different sets of communalities, all of which reduce the matrix of correlations of our four tests to rank 2. If, for example, we fix the first communality arbitrarily, say at ·5, we can condense the determinant to one of order 3 by using ·5 as a pivot (as on page 22) except that the diagonal of the smaller matrix will be blank:

(·5)	·4	·4	·2
·4	.	·7	·3
·4	·7	.	·3
·2	·3	·3	.
	.	·19	·07
	·19	.	·07
	·07	·07	.

We can then fill the diagonal of the smaller matrix with numbers which will make each of its tetrads zero, namely—

.	·19	·19	·0258

and then, working back to the original matrix, find the communalities—

·5	·7	·7	·1316

which make its rank exactly 2. We can similarly insert different numbers for the first communality and calculate different sets of communalities, any one set of which will reduce the rank to 2. In this way we can go from 1·0 down to 0·22951 for the first communality without obtaining inadmissible magnitudes for the others. Some sets are given in the following table * :

1	2	3	4	*Sum*
1·0	·7	·7	·12963	2·52963
·7	·7	·7	·13030	2·23030
·5	·7	·7	·13158	2·03158
·3	·7	·7	·14	1·84
·2$\dot{6}$	·7	·7	·15	1·81$\dot{6}$
·256	·7	·7	·158$\dot{3}$	1·814$\dot{3}$
·25	·7	·7	·1$\dot{6}$	1·81$\dot{6}$
·24	·7	·7	·20	1·84
·23	·7	·7	·7	2·33
·22951	·7	·7	1·0	2·62951

If, however, we search for and find a fifth test to add to the four, which will still permit the rank to be reduced to 2, this fifth test will fix the communalities at some point or other within the above range. Suppose that this test gave the correlations shown in the last row and column :

	1	2	3	4	5
1	.	·4	·4	·2	·5883
2	·4	.	·7	·3	·2852
3	·4	·7	.	·3	·2852
4	·2	·3	·3	.	·1480
5	·5883	·2852	·2852	·1480	.

If we now try to find communalities to reduce this matrix to rank 2 (as can be done), we find only the one set—

·7 ·7 ·7 ·13030 ·5

The reader can try this by assigning an arbitrary value for

* The circumstance that the communalities of Tests 2 and 3 remain fixed and alike is due to these tests being identical except for their specific. This lightens the arithmetic, but would not occur in practice.

the first one,* and then condensing the matrix on the lines employed above, when he will always find some obstacle in the way unless he chooses ·7. Try, for example, ·5 for the first communality:

(·5)	·4	·4	·2	·5883
·4	.	·7	·3	·2852
·4	·7	.	·3	·2852
·2	·3	·3	.	·1480
·5883	·2852	·2852	·1480	.

(x)	·27	·07	− ·09272
·27	.	·07	− ·09272
·07	·07	.	− ·04366
− ·09272	− ·09272	− ·04366	.

Now, if the upper matrix is to be of rank 2, the second condensation must give only zeros (see footnote, page 22). But if we fix our attention on different tetrads in the lower matrix which contain the pivot x, we see that they give, if they have to be zero, incompatible values for x. Thus from one tetrad we get $x = ·27$, from another $x = ·14866$. With ·5 as first communality, rank 2 cannot be attained. With five tests (or more), if rank 2 can be attained at all, it can only be by one unique set of communalities. Just as it took three tests to enable the saturations with Spearman's g to be calculated, so it takes five tests to enable communalities due to two common factors to be calculated. For larger numbers of common factors, the number of tests required to make the set of communalities unique is shown in the following table (*Vectors*, 77). The lower numbers are given by the formula—

$$n \geqslant \frac{(2r + 1) + \sqrt{(8r + 1)}}{2}$$

r Factors	1	2	3	4	5	6	7	8	9	10	11	12
n Tests	3	5	6	8	9	10	12	13	14	15	17	18

* Alternatively, the communalities (which are now unique) can be found by equating to zero those three-rowed minors which have only one element in common with the diagonal (*Vectors*, 86). In this connection see Ledermann, 1937.

If we were actually confronted with the matrix of correlations shown on page 39, and asked what the communalities were which reduced it to the lowest possible rank, we would find it very unsatisfactory to have to guess at random and try each set; and our embarrassment would be still greater if there were more tests in the battery, as would actually be the case in practice. There would also be sampling error (which in this our preliminary description of Thurstone's method we are assuming to be non-existent). Under these circumstances, devices for arriving rapidly at approximate values of the communalities are very desirable. The plan adopted by Thurstone will be described in Chapter X.

CHAPTER III

THE SAMPLING THEORY

1. *Two views. A hierarchical example as explained by one general factor.*—The advance of the science of factorial analysis of the mind to its present position has not taken place without opposition, and it is the purpose of the present chapter to give a preliminary description of some objections which have been frequently raised by the present writer (Thomson, 1916, 1919*a*, 1935*b*, etc.) and which indeed he still holds to, although there has been of late years a considerable change of emphasis in the interpretations placed upon factors by the factorists themselves, which have tended to remove his objections. Briefly, the opposition between the two points of view would disappear if factors were admitted to be only statistical coefficients, possibly without any more " reality " than an average, or an index of the cost of living, or a standard deviation, or a correlation coefficient—though, on the other hand, it may be admitted that some of them, Spearman's *g* for example, may come to have a very real existence in the sense of being both useful and influential in the lives of men.

There seems to be room for some form of integration of a number of apparently antithetical ideas regarding the way in which the mind functions, and the sampling theory which the writer has put forward * seems in particular to show that what have been called " monarchic," " oligarchic," and " anarchic " doctrines of the mind (*Abilities*, Chapters II–V) are very probably only different ways of describing the same phenomena.

The contrast—perhaps one should say the apparent

* For a general statement see Brown and Thomson, 1921, Chapter X, and Thomson, 1935*b*, and references there given. A somewhat similar point of view has in more recent years been taken in America by R. C. Tryon, 1932*a* and *b*, and 1935.

contrast—between the factorial and the sampling points of view * can be best seen by considering the explanation of the same set of correlation coefficients by both views. As we have consistently done, so far, in this part of our book, we shall again suppose that there are no experimental or sampling errors—we shall consider them abundantly in due course—and to simplify the argument we shall take in the first place a set of correlation coefficients whose tetrads are exactly zero, which can therefore be completely " explained " by a general factor *g* and specifics, as in this table :

	1	2	3	4
1	.	·746	·646	·527
2	·746	.	·577	·471
3	·646	·577	.	·408
4	·527	·471	·408	.

We can more exactly follow the argument if we employ the vulgar fractions of which these are the decimal equivalents, namely the following, each divided by 6 :

	1	2	3	4
1	.	$\sqrt{20}$	$\sqrt{15}$	$\sqrt{10}$
2	$\sqrt{20}$	.	$\sqrt{12}$	$\sqrt{8}$
3	$\sqrt{15}$	$\sqrt{12}$	.	$\sqrt{6}$
4	$\sqrt{10}$	$\sqrt{8}$	$\sqrt{6}$	.

In this form the tetrad-differences are all obviously zero by inspection. These correlations can therefore be explained by one general factor, as in Figure 10, which gives them exactly.

We have here a general factor of variance 30 which is the sole cause of the correlations, and specific factors of variances 6, 15, 30, and 60. The variances of the four

* Two papers by S. C. Dodd (1928 and 1929) gave a very full and competent comparison of the two theories up to that date. The present writer agrees with a great deal, though not with all, of what Dodd says ; but see the later paper (Thomson, 1935*b*) and also Chapter XVIII of this book.

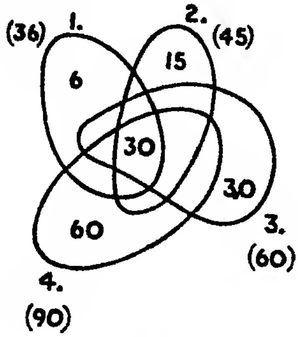

Figure 10.

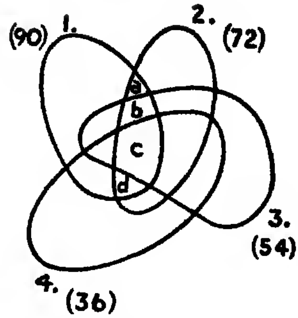

Figure 11.

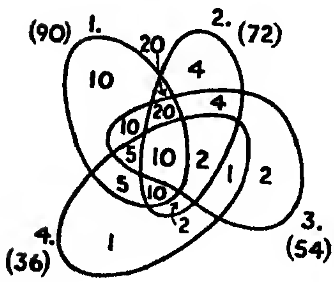

Figure 12.

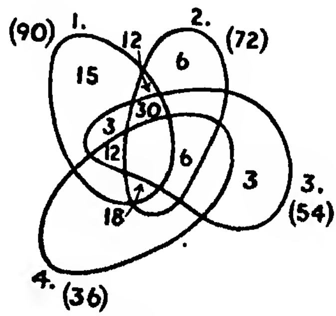

Figure 13.

" tests " are 36, 45, 60, and 90. The " communalities " and " specificities " are :

Test	1	2	3	4	*Totals*
Communality	$\frac{30}{36}$	$\frac{30}{45}$	$\frac{30}{60}$	$\frac{30}{90}$	$\frac{420}{180} = 2{\cdot}333$
Specificity	$\frac{6}{36}$	$\frac{15}{45}$	$\frac{30}{60}$	$\frac{60}{90}$	$\frac{300}{180} = 1{\cdot}667$
Totals	1	1	1	1	4

These communalities can be calculated from the correlation coefficients, for it will be remembered (Chapter I, Section 4) that when tetrad-differences are exactly zero, each correlation coefficient can be expressed as the

product of two correlation coefficients with g (two "saturations"). Thus—

$$r_{12} = r_{1g} r_{2g}$$
$$r_{13} = r_{1g} r_{3g}$$
$$r_{23} = r_{2g} r_{3g}$$

Therefore—

$$\frac{r_{12} r_{13}}{r_{23}} = \frac{(r_{1g} r_{2g})(r_{1g} r_{3g})}{(r_{2g} r_{3g})} = r_{1g}^2$$

the square of the saturation of Test 1 with g. And when there is only one common factor, the square of its saturation is the communality.

The quantity $r_{12} r_{13}/r_{23}$, therefore, means, on this theory of one common factor, the communality, or square of the saturation with g, of the first test. Its value in our example is 30/36, or five-sixths.

2. *The alternative explanation. The sampling theory.*—The alternative theory to explain the zero tetrad-differences is that each test calls upon a *sample of the bonds* which the mind can form, and that some of these bonds are common to two tests and cause their correlation. In the present instance we have arranged this artificial example so that the tests can be looked upon as samples of a very simple mind, which can form in all 108 bonds (or some multiple of 108).* The first test uses five-sixths of these (or 90), the second test four-sixths (or 72), the third three-sixths (54), and the fourth two-sixths (or 36). These fractions are the same in value as the communalities of the former theory. Each of them may be called the "richness" of the test. Thus Test 1 is most rich, and draws upon five-sixths of the whole mind. The fractions $r_{ij} r_{ik}/r_{jk}$, which in the former theory were "communalities," are in the sampling theory "coefficients of richness." They formerly indicated the fraction of each test's variance supplied by g; they indicate here the fraction which each test forms of the whole "mind" (but see later, concerning "sub-pools").

* There is nothing mysterious about the number 108. It is chosen merely because it leads to no fractions in the diagram. Any large number would do.

Now, if our four tests use respectively 90, 72, 54, and 36 of the available bonds of the mind, as indicated in Figure 11, then there may be almost any kind of overlap between two of the tests. Any of the cells of the diagram may have contents, instead of all being empty except for g and the specifics. If we know nothing more about the tests except the fractions we have called their "richnesses," we cannot tell with certainty what the contents of each cell will be; but we can calculate what the *most probable* contents will be. If the first test uses five-sixths and the second test four-sixths of the mind's bonds, it is most probable that there will be a number of bonds common to both tests equal to $\frac{5}{6} \times \frac{4}{6}$, or 20/36ths of the total number. That is, the four cells marked a, b, c, d in the diagram, the cells common to Tests 1 and 2, will most likely contain—

$$\frac{20}{36} \times 108 = 60 \text{ bonds}$$

between them. By an extension of the same principle we can find the most probable number in each cell. Thus c, the number of bonds used in all four of the tests, is most probably—

$$\frac{5}{6} \times \frac{4}{6} \times \frac{3}{6} \times \frac{2}{6} \times 108 = 10 \text{ bonds.}$$

In this way we reach the most probable pattern of overlap of the four tests shown in Figure 12. *And this diagram gives exactly the same correlations as did Figure* 10. Let us try, for example, the value of r_{23} in each diagram. In Figure 10 we had—

$$r_{23} = \frac{30}{\sqrt{(45 \times 60)}} = \frac{\sqrt{12}}{6} = \cdot 577$$

In Figure 12 the same correlation is—

$$r_{23} = \frac{20 + 10 + 4 + 2}{\sqrt{(72 \times 54)}} = \frac{\sqrt{12}}{6} = \cdot 577$$

This form of overlap, therefore, will give zero tetrad-differences, just as the theory of one general factor did. More exactly, this sampling theory gives zero tetrad-

differences as the *most probable* (though not the certain) connexion to be found between correlation coefficients (Thomson, 1919*a*).

If we let p_1, p_2, p_3, and p_4 represent fractions which the four tests form of the whole pool of N bonds of the mind, then the number common to the first two tests will most probably be p_1p_2N, and the correlation between the tests

$$r_{12} = \frac{p_1p_2N}{\sqrt{(p_1N \cdot p_2N)}} = \sqrt{p_1p_2}$$

We therefore have, in any tetrad, quantities like the following:

	3	4
1	$\sqrt{p_1p_3}$	$\sqrt{p_1p_4}$
2	$\sqrt{p_2p_3}$	$\sqrt{p_2p_4}$

and the tetrad-difference is, most probably (Thomson, 1927*a*, 253)—

$$\sqrt{p_1p_3p_2p_4} - \sqrt{p_1p_4p_2p_3} = 0$$

This may be expressed by saying that the laws of probability alone will cause a tendency to zero tetrad-differences among correlation coefficients. In another form, which will be useful later, this statement can be worded thus: The laws of probability or chance cause any matrix of correlation coefficients to tend to have rank 1, or at least to tend to have a low rank (where by rank we mean the maximum order among those non-vanishing minors which avoid the principal diagonal elements).

It is, in the opinion of the present writer, this fact—a result of the laws of chance and not of any psychological laws—which has made conceivable the analysis of mental abilities into a few common factors (if not into one only, as Spearman hoped) and specifics. Because of the laws of chance the mind works *as if* it were composed of these hypothetical factors *g*, *v*, *n*, etc., and a number of specific factors. The causes may be "anarchic," meaning that they are numerous and unconnected, yet the result is "monarchic," or at least "oligarchic," in the sense that it may be so described—provided always that large specific factors are allowed.

Of course, if the tetrad-differences actually found among correlation coefficients of mental tests were really exactly zero, or so near to zero that the discrepancies could be looked upon as "errors" due to our having tested a particular set of persons who did not accurately represent the whole population, then the theory of only one general factor would have to be accepted. For it gives exactly zero tetrad-differences, whereas the sampling theory only gives a tendency in that direction. But in actual fact it is only a tendency which is found, and matrices of correlation coefficients do not give zero tetrad-differences until they have been carefully purified by the removal of tests which "break the hierarchy." It has not proved very difficult to arrive at such purified teams of hierarchical tests. That is to be expected on the Sampling Theory, according to which hierarchical order is the *most probable* order. In the same way one would not have to go on throwing ten pennies for long before arriving at a set which gave five heads and five tails, for that is the most probable (yet not the certain) result.

8. *Specific factors maximized.*—The specific factors play, in the Spearman and Thurstone methods of factorization, an important rôle, and our present example can be used to illustrate the fact, which is not usually realized, that both these methods *maximize the specifics* (Thomson, 1938*c*) by their insistence on *minimizing* the number of general factors. In Figure 10, of the whole variance of 4, the specific factors contribute 1·667, or 41·7 per cent. In Figure 12, they contribute only—

$$\frac{10}{90} + \frac{4}{72} + \frac{2}{54} + \frac{1}{36} = \frac{250}{1{,}080} = \cdot 2315, \text{ or } 5\cdot 8 \text{ per cent.}$$

Apart from certain trivial exceptions which do not occur in practice, it is generally true that minimizing the number of common factors maximizes the variance of the specifics. Numerous other analyses of the above correlations can be made (Thomson, 1935*c*), but they all give a variance to the specifics which is less than 1·667. Here, for example, in Figure 13 (page 44), is an analysis which has no general

factor but six other common factors, and which gives a total specific variance of—

$$\frac{15}{90} + \frac{6}{72} + \frac{3}{54} + 0 = \frac{330}{1,080} = \cdot 3056, \text{ or } 7\cdot 6 \text{ per cent.}$$

The same principle, that reducing the number of common factors tends to increase the variance of the specifics, can be seen illustrated in Figures 5 and 6 (Chapter I, page 15). Figure 6 has five common factors, and the proportion which the specific variance bears to the whole four tests is—

$$\frac{1}{10} + \frac{1}{10} + \frac{1}{10} + \frac{1}{10} = 0\cdot 4, \text{ or } 10 \text{ per cent.}$$

In Figure 5 there are only two common factors, and the specific variance has risen to—

$$\frac{5}{10} + \frac{2}{10} + \frac{2}{10} + \frac{5}{10} = 1\cdot 4, \text{ or } 35 \text{ per cent.}$$

Again, in Figures 7, 8, and 9 (Chapter II, pages 26, 34, and 38) the same phenomenon can be observed. In Figure 7, with three common factors, the specific variances form 37·5 per cent. of the four tests; in Figures 8 and 9, with only two common factors, the specific variances form 54·6 per cent.

Now, specific factors are undoubtedly a difficulty in any analysis, and to have the specific factors made as large and important as possible is a heavy price to pay for having as few common factors as possible.

Spearman, it is true, in his earlier writings, and in Chapter IX of *The Abilities of Man*, boldly accepts the idea of specific factors; that is, factors which play no part except in one activity only, or in very closely allied activities. His analogy of "mental energy" (*g*) and "neural machines" (the specifics) always makes a considerable appeal to an audience. On that analogy the energy of the mind is applicable in any of our activities, as the electric energy which comes into a house is applicable in several different ways: in a lighting-bulb, a radio set, a cooking-stove, a heater, possibly an electric razor, etc. Some of the specific machines which use the electric energy need

more of it than do others, just as some mental activities are more highly saturated with *g*. If it fails, they all cease to work; if it weakens, they all work badly. Yet when it is strong, they do not all work equally well: the electric carpet-sweeper may function badly while the electric heater functions well, because of a faulty connection in the (specific) carpet-sweeping machine; while Jones next door (enjoying the same general electric supply) possesses no electric carpet-sweeper. So two men may have the same *g*, but only one of them possess the specific neural machine which will enable him to perform a certain mental task. The analogy is attractive, and, it must be agreed, educationally and socially useful. There is no objection to accepting it so far. But with the complication of group factors it begins to break down. Most activities are found to require the simultaneous use of several "machines." There does not seem so sharp a distinction between the machines and the general energy. Moreover, the general energy, if there be such a thing, of our personalities is commonly held to be of instinctive and emotional nature rather than intellective, while *g*, whatever else it is, is commonly thought of as closely connected with intelligence.

That specific factors are a difficulty seems to be recognized by Thurstone. "The specific variance of a test," he writes (*Vectors*, 63), "should be regarded as a challenge," and he looks forward to splitting a specific factor up into group factors by brigading the test in question with new companion tests in a new battery. It seems clear that the dissolution of specifics into common factors is unlikely to happen if each analysis is conducted on the principle of making the specific variances as large as possible. We must, however, leave this point here, to return to it in a later chapter of this book.

4. *Sub-pools of the mind.*—A difficulty which will occur to the reader in connexion with the sampling theory is that, when the correlation between two tests is large, it seems to imply that each needs nearly the whole mind to perform it (Spearman, 1928, 257). In our example the correlation between Tests 1 and 2 was ·746, a correlation not infre-

quently reached between actual tests. It is, for instance, almost exactly the correlation reported by Alexander between the Stanford-Binet test and the Otis Self-administering test (Alexander, 1935, Table XVI). Does this, then, mean that each of these tests requires the activity of about four-sixths or five-sixths of all the "bonds" of the brain? Not necessarily, even on the sampling theory. These two tests are not so very unlike one another, and may fairly be described as sampling the same region of the mind rather than the whole mind, so that they may well include a rather large proportion of the bonds found in that region. They may be drawn, that is, from a sub-pool of the mind's bonds rather than from the whole pool (Thomson, 1935*b*, 91; Bartlett, 1937*a*, 102). Nor need the phrase "region of the mind" necessarily mean a topographical region, a part of the mind in the same sense as Yorkshire is part of England. It may mean something, by analogy, more like the lowlands of England, all the land easily accessible to everybody, lying below, say, the 300-foot contour line. What the "bonds" of the mind are, we do not know. But they are fairly certainly associated with the neurones or nerve cells of our brains, of which there are approaching one hundred thousand million in each normal brain. Thinking is accompanied by the excitation of these neurones in patterns. The simplest patterns are instinctive, more complex ones acquired. Intelligence is possibly associated with the number and complexity of the patterns which the brain can (or could) make. A "region of the mind" in the above paragraph may be the domain of patterns below a certain complexity, as the lowlands of England are below a certain contour line. Intelligence tests do not call upon brain patterns of a high degree of complexity, for these are always associated with acquired material and with the educational environment, and intelligence tests wish to avoid testing acquirement. It is not difficult to imagine that the items of the Stanford-Binet test call into some sort of activity nearly all the neurones of the brain, though they need not thereby be calling upon all the patterns which those neurones can form. When a teacher is

demonstrating to an advanced class that "a quadratic form of rank 2 is identically equal to the product of two linear forms," he is using patterns of a complexity far greater than any used in answering the Binet-Simon items. But the neurones which form these patterns may not be more numerous. Those complicated patterns, however, are forbidden to the intelligence tester, for a very intelligent man may not have the ghost of an idea what a "quadratic form" is. Within the limits of the comparatively simple patterns of the brain which they evoke, it seems very possible that the two tests in question call upon a large proportion of these, and have a large number in common. The hope of the intelligence tester is that two brains which differ in their ability to form readily and clearly the comparatively simple patterns required by his test will differ in much the same way if, given the same educational and vocational environment, they are later called upon to form the much more complex patterns there found.

As has been indicated, the author is of opinion that the way in which they magnify specific factors is the weak side of the theories of a single general factor or of a few common factors. That does not mean, however, that a description of a matrix of correlations in terms of these theories is inexact. Men undoubtedly do perform mental tasks *as if* they were doing so by means of a comparatively small number of group factors of wide extent, and an enormous number of specific factors of very narrow range but of great importance each within its range. Whether a description of their powers in terms of the few common factors only is a good description depends in large measure on what purpose we want the description to subserve. The practical purpose is usually to give vocational or educational advice to the man or to his employers or teachers, and a discussion of the relative virtues of different theories in this respect must wait until we have considered the somewhat technical matter of "estimation" in later chapters. We shall there see that factors, though they cannot improve and indeed may blur the accuracy of vocational estimates, may, however, facilitate them where otherwise they would have been

impossible, as money facilitates trade where barter is impossible.

As a theoretical account of each man's mind, however, the theories which use the smallest number of common factors seem to have drawbacks. They can give an exact reproduction of the correlation coefficients. But, because of their large specific factors, they do not enable us to give an exact reproduction of each man's scores in the original tests, so that much information is being lost by their use. Reproduction of the original scores with complete exactitude can only be achieved by using as many factors as there are tests. But it can be done with considerable accuracy by a few of Hotelling's factors (called " principal components "), which will be described later.

It will be seen from considerations such as these that alternative analyses of a matrix of correlations, even although they may each reproduce the correlation coefficients exactly, may not be equally acceptable on other grounds. The sampling theory, and the single general factor theory, can both describe exactly a hierarchical set of correlation coefficients, and they both give an explanation of why approximately hierarchical sets are found in practice. In a mathematical sense, they are alternatives. But as Mackie has shown (Mackie, 1928*b*), a psychologist who believes that the " bonds " of the sampling theory have any real existence, in the sense, say, of being represented in the physical world by chains and patterns of neurones, cannot without absurdity believe in the similarly real existence of specific factors. The analogue to Spearman's *g*, on the sampling theory, is simply the whole mind. " How, then," (as Mackie asks) " can we have other factors independent of such a factor as this ? " Only by the formal device of letting the specific factor include the annulling of the work done by the other part of the mind, a legitimate mathematical procedure but not one compatible with actual realities. Either, then, we must give up the factors of the two-factor theory, or the bonds of the sampling theory, as realities. We cannot keep both as realities, though we may employ either mathematically.

5. *The inequality of men.*—Professor Spearman has

opposed the sampling theory chiefly on the ground that it would make all correlations equal (and zero), and involve the further consequence that all men are equal in their average attainments (*Abilities*, 96), if the number of elementary bonds is large, as the sampling theory requires. Both these objections, however, arise from a misunderstanding of the sampling theory, in which a sample means "some but not all" of the elementary bonds (Thomson, 1935*b*, 72, 76). As has been explained, tests can differ, on this theory, in their richness or complexity, and less rich tests will tend to have low, more complex tests will tend to have high correlations, at any rate if the "bonds" tend to be *all-or-none* in their nature, as the action of neurones is known to be. Neurones, like cartridges, either fire or they don't. And as for the assertion that the theory makes all men equal, there is no basis whatever for the suggestion that it assumes every man to have an equal chance of possessing every element or bond. On the contrary, the sampling theory would consider men also to be samples, each man possessing some, but not all, both of the inherited and the acquired neural bonds which are the physical side of thought. Like the tests, some men are rich, others poor, in these bonds. Some are richly endowed by heredity, some by opportunity and education; some by both, some by neither. The idea that men are samples of all that might be, and that any task samples the powers which an individual man possesses, does not for a moment carry with it the consequences asserted of equal correlations and a humdrum mediocrity among human kind.

CHAPTER IV

THE GEOMETRICAL PICTURE

1. *The fundamental idea.*—The student reading articles on factorial analysis is continually coming across geometrical and spatial expressions which he may be surprised to find in a psychological setting. For example, in Section 8 of our Chapter II we spoke of "rotating" the loadings of Thurstone's "centroid" method until they fulfil certain conditions. These geometrical expressions arise from the fact that the mathematics of mental testing is the same in its formal aspect as the mathematics of multi-dimensional space, and it is the object of the present chapter to explain this in elementary terms. Some degree of understanding of this is essential for the worker with tests, and it is not difficult when divested as far as possible of the algebraic symbols in which it is usually clothed.

The fundamental idea is that the correlation between two tests can be pictorially represented by the angle between two lines which stand for the two tests, and which pass through a point, thus forming an X with its legs stretching ever so far in both ways. The point where the lines cross represents a man who has the average score on both tests. Other points on the lines represent standardized scores in the tests which are more or less removed from the average—an arrowhead can be placed on each line to represent the positive direction, as in Figure 14. If the lines—taken in the direction of these arrowheads—make only a small angle with one another, they represent tests which are highly correlated. As the correlation decreases, this angle increases. When the correlation is zero, the angle

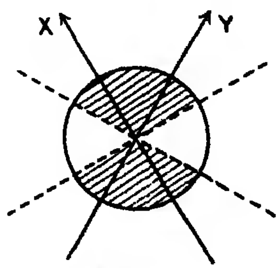

Figure 14.

is a right angle. If the angle becomes obtuse, the correlation is negative.

Any point on the paper then represents a person by his two standardized scores in these two tests, obtained by dropping perpendiculars on to the two lines representing the tests. If we were to measure a large number of persons by each of these two tests—say, ten thousand persons—and place a dot on the paper for each person as represented by his two scores, we would naturally find that these dots would be crowded most closely together round the point where the test lines (or test *vectors*, as they are technically called) cross, where the average man is situated. The ten thousand dots would look, in fact, like shot marks on a target of which the bull's-eye was the average man at the cross-roads of the test vectors. The density of the dots would fall off equally to the north, south, east, and west of this point. Their "contours of density," as we say, would be circles. Circles, because *any* line through the imaginary man-who-is-average-in-everything represents a conceivable test, and the standard deviation is everywhere represented by the same unit of length. The dots would look exactly like a crowd which, equally in all directions, was surrounding a focus of attraction at the crossing-point of the tests.

2. *Sectors of the crowd.*—On the diagram are shown also two dotted lines, perpendicular respectively to the two test vectors. Persons who are standing on one of these dotted lines have exactly the average score in the test to which it is perpendicular. Two of the sectors of the crowd are distinguished by shading in the diagram. Let us fix our attention on the northern shaded sector, which includes the two positive directions of the test vectors, marked by the arrowheads. Everybody in this sector of the crowd has a score above the average in both tests. Similarly, in the other shaded sector of the crowd, everybody has a score below the average in both tests. Both these sectors of the crowd contribute to the correlation between the tests, since everybody in these sectors does well in both, or badly in both.

The people in the white sectors of the crowd, however,

have scores above the average in one test and below the average in the other. They diminish the correlation between the tests. Those in the western white sector have scores above the average in Test X, but below the average in Test Y; and vice versa for those in the eastern white sector.

If the arrowheads X and Y are brought nearer together (while the people in the circular crowd remain standing still), so that the angle between the test vectors is diminished, the dotted lines will move so as to diminish the white sectors which lie between them, and the correlation will increase. When the test vectors are close together, one coinciding with the other, the white sectors will have disappeared and the correlation will be perfect. When the test vectors are at right angles, the white sectors will be quadrants, the crowd will be half "black" and half "white," and the correlation zero. Beyond the right-angle position, there will be more white than black, and a negative correlation.

It is clear, then, that the angle between the test vectors inversely represents the correlation between the tests. It can be shown (but we shall take it on trust) that the *cosine* of the angle is equal to the correlation (Garnett, 1919*a*; Wilson, 1928*a*). If we wish, therefore, to draw two vectors for two tests whose correlation we know, we consult a table of trigonometrical ratios, to find the angle whose cosine is equal to the correlation coefficient, and draw the lines accordingly.

8. *A third test added. The tripod.*—If we now wish to draw the vector of a third test, we must similarly consult the trigonometrical table to find from its correlation coefficients the angles it makes with the two former tests. We shall then usually discover that we cannot draw it on our paper, but that it has to stick out into a third dimension. It will only lie in the same plane as the other two if either the sum or the difference of its angles with them equals the angle between the first two tests. Usually this will not be the case, and the vectors of these tests will require three-dimensional space. They will look like a tripod extended upwards as well as downwards. If the correla-

tions are high, the tripod's legs will be close together; if low, they will be far apart. This tripod analogy will make plausible to the reader the assertion that some sets of correlation coefficients cannot logically coexist. For the legs of a tripod cannot take up positions at *any* angles. If two of the angles are very small, the third one cannot be very large. The sum of any two of the angles must at least equal the third angle. And so on. For example, the following matrix of correlations is an impossibility :

	1	2	3
1	1·00	·34	·77
2	·34	1·00	·94
3	·77	·94	1·00

Here Tests 1 and 2 are highly correlated with Test 3, so highly that they cannot possibly have only a correlation coefficient of ·34 with each other. The angles corresponding to the above coefficients (taken as cosines) are :

	1	2	3
1	0°	70°	40°
2	70°	0°	20°
3	40°	20°	0°

and the fact that 40° + 20° is less than 70° shows that the matrix is impossible.

When the symmetrical matrix of correlations is an impossible one, which could not really occur, it will be found that either the determinant itself, or one of the pivots in the calculation explained in Chapter II, Section 2, is *negative*. Let us carry out the calculation for the above matrix :

(1·00)	·34	·77
·34	1·00	·94
·77	·94	1·00
	(·8844)	·6782
	·6782	·4071

Determinant = —·0999

This test serves also for larger matrices.

Let us, however, return to our tripod of three vectors which by their angles with one another represent the correlations of three tests—the legs of the tripod being the negative directions of the tests, let us assume, and their continuation upward past their common crossing-point the positive directions, though this is not essential.

The point where the three vectors cross represents the average man, who obtains the average score (which we will agree to call zero) on each of the three tests. Any other point in space represents a man whose scores in the three tests are given by the feet of perpendiculars from this point on to the three test vectors. If, again, we suppose that ten thousand persons have undergone these three tests, the space round the test vectors will be filled with ten thousand points, which will be most closely crowded together near the average man at the crossing-point (or "*origin*") of the vectors, and will form a spherical swarm falling off in density equally in all directions from that point.

4. *A fourth test added.*—One test was represented by a line. Two tests by two lines in a plane. Three tests by three lines in ordinary space. Suppose now we have a fourth test, look up its angles with the pre-existing three tests, and try to draw its line or vector, adding a fourth leg to the tripod. Just as the third test would not usually lie in the plane of the first two, but required a third dimension to project out into, so the fourth test will not usually be capable of being represented in the three-space of the three tests. Its angles with them will not fit unless we add a *fourth* dimension.

Here, of course, the geometrical picture, strictly speaking, breaks down. But it is usual and mathematically helpful to continue to speak as though spaces of higher dimensions really existed. In a "space" of four dimensions we can imagine four test vectors crossing at a point, their angles with one another depending upon the correlations. We can imagine a "spherical" swarm of dots representing persons. And when we add more tests, we can similarly imagine spaces of 5, 6 . . . *n* dimensions to accommodate their test vectors. The reader should not allow the im-

possibility of *visualizing* these spaces of higher dimensions to trouble him overmuch. They are only useful forms of speech, useful because they enable us to refer concisely to operations in several variables which are exactly analogous to familiar operations in the real space in which we live—such as " rotating " a line or a set of lines round a pivot.

5. *Two principal components.*—Let us now express the ideas we have used in the preceding three chapters in terms of this geometrical picture. Independent factors will be represented by vectors at right angles to one another (we shall for the most part be concerned only with independent, i.e. uncorrelated factors, though at a later stage we shall have something to say about correlated or " oblique " factors). Analysing a set of tests into independent factors means, in terms of our geometrical picture, referring their test vectors to a set of rectangular vectors as axes of co-ordinates—the Greek equivalent " orthogonal " is generally used in this connexion instead of "rectangular." Let us explain this first of all in the simplest case, that of two tests, represented by their vectors in a plane, at the angle corresponding to their correlation.

In this case, the most natural way of drawing orthogonal co-ordinates on the paper is to place one of them (see Figure 15) half-way between the test vectors, and the other, of course, at right angles to the first. These factor vectors correspond, in fact, to Hotelling's "principal components," to which we shall return later. Of these two factors (or components) OA is as near as it can be to both test vectors—it is the "first principal component."

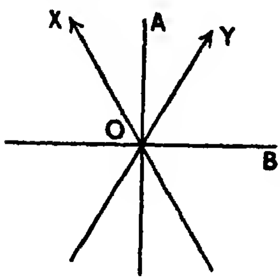

Figure 15.

We pictured, before, a swarm of ten thousand dots on the paper, each representing a person by his scores in the two tests, found by dropping perpendiculars from his dot to the two vectors. Instead of describing each point (each person, that is) by the two test scores, it is clear that we

could describe it by the two factor scores—using rectangular instead of oblique co-ordinates. It is also clear that, as far as this purpose goes, we might have taken our factor vectors or factor axes anywhere, and not necessarily in the positions *OA* and *OB*, provided they went through the point *O* and were at right angles. In other words, we can "rotate" *OA* and *OB* round the point *O*, and any position is equally good for describing the crowd of persons. Either of the tests, indeed, might be made one of the factors. The positions shown in Figure 15 are advantageous only if we want to use only one of our factors and discard the other, in which case obviously *OA* is the one to keep, as it lies as near as possible to both test vectors.* The scores along *OA* are the best possible single description of the two test results. That is the distinguishing virtue of Hotelling's "first principal component."

6. *Spearman axes for two tests.*—The orthogonal axes chosen by Spearman for his factors are, however, none of the positions to which *OA* and *OB* can be rotated in the plane of the paper. Besides, Spearman has three factors, and therefore three axes, for two tests, namely the general factor and the two specific factors, and we cannot have three orthogonal axes or factor vectors on a sheet of paper. The Spearman factors must, for two tests, lie in three-dimensional space, like the three lines which meet in the corner of a room. If we rotate the *OA* and *OB* of Figure 15 out of the plane of the paper (say, pushing *A* below the surface of the paper, and, say, raising *B* above it), we shall clearly have to add a third axis, at right angles to *OA* and *OB*, to enable us to describe the tests and the persons who remain on the paper. There are now three axes to rotate; and they must rotate rigidly, remaining at right angles to one another. The point at which Spearman stops the rotation, and decides that the lines then represent the "best" factors, is a position in which one of the axes is

* Persons will, in fact, be placed in the same order of merit by their factors *A* as they are placed in by their *average* scores on the two tests, but this is not the case with the Hotelling first component of larger numbers of tests.

at right angles to Test X, and another is at right angles to Test Y. The third axis then represents g.

7. *Spearman axes for four tests.*—We are accustomed to depicting three dimensions on a flat sheet of paper, and so we can, in Figure 16, represent the Spearman axes g, s_1, and s_2 for two tests. And since we have begun to depict other dimensions, by means of perspective, on a flat sheet, let us continue the process and by a kind of super-perspective imagine that the lines s_3, s_4, and any others we may care to add, represent axes sticking out into a fourth, a fifth, and higher dimensions. Figure 16 thus represents the five Spearman axes for four tests, of which only the vector of the first test is shown (in its positive half only).

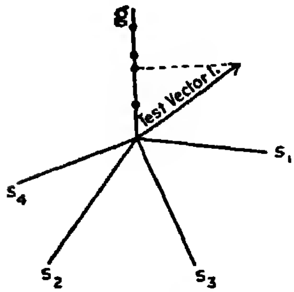

Figure 16.

All the five lines g, s_1, s_2, s_3, and s_4 must be imagined as being each at right angles to all the others in five-dimensional space. The vector of Test 1, shown in the diagram, lies *in* the plane or wall edged by g and s_1. It forms acute angles with g and with s_1, the cosines of which angles are its saturations with g and s_1 respectively. If it had been highly saturated with g, it would have leaned nearer to g and farther away from s_1.

The other three axes, s_2, s_3, and s_4, are all at right angles to the wall or plane in which Test 1 lies. They have, therefore, no correlation with Test 1, no share in its composition. Test vector 2 similarly lies in the wall edged by g and s_2, test vector 3 in that edged by g and s_3. The axis g forms a common edge to all these planes. If the battery of tests is hierarchical—that is, if the tetrad-differences are all zero—then all the tests of the battery can be depicted in this way, each in its own plane at right angles to all the other planes, no test vector being in the spaces between the "walls."

The four test vectors themselves, of course, are only in a four-dimensional space (a 4-space we shall say, for

brevity). Just as, when we were discussing Figure 15, we said that Spearman used three axes which were all out of the plane of the paper, so here in Figure 16, with four test vectors (only one shown) in a 4-space, Spearman uses five axes in a space of one dimension higher than the number of tests. For n hierarchical tests, Spearman's factors are in an $(n + 1)$-space.

If along each test vector we measure the same distance as a unit, then perpendiculars from these points on to the g axis will give the saturations of the tests with g as fractions of this unit distance. The four dots on the g axis in Figure 16 may thus be taken as representing the test vectors projected on to the "common-factor space," which is here a line, a space of one dimension only. Thurstone's system is like Spearman's except that the common-factor space is of more dimensions, as many as there are common factors. Figure 17 shows the Thurstone axes for four tests whose matrix of correlation coefficients can be reduced to rank 2.

8. *A common-factor space of two dimensions.*—Here there are two common factors, a and b, and four specifics, s_1, s_2, s_3, and s_4. All the six axes representing these factors in the figure are to be imagined as existing in a 6-space, each at right angles to all the others. The common-factor space is here two-dimensional, the plane or wall edged by a and b—to make it stand out in the figure, a door and a window have been sketched upon it.

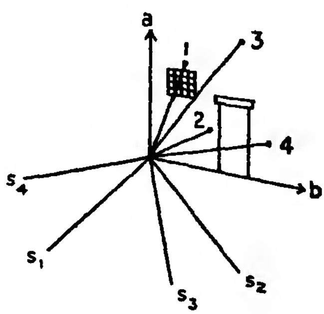

Figure 17.

In Spearman's Figure 16, each test vector lay in a plane defined by g and one of the specific axes. Here in Figure 17, each test vector lies in a different 3-space. These different 3-spaces have nothing in common with one another except the plane ab, the wall with the door and window in the digram. In Figure 16 the projections of the test vectors on to the common-factor space were lines which all coincided in direction (though they were of different lengths), for

there the common-factor space was a line. Here the common-factor space is a plane, and the projections of the four test vectors on to that plane are shown in the figure by the lines on the "wall." These lines, if they are all projections of vectors of unit length, will by their lengths on the wall represent the square roots of the communalities.

9. *The common-factor space in general.*—When there are r common factors, the common-factor space is of r dimensions, and the whole factor space (including the specifics) is of $(n + r)$ dimensions. The test vectors themselves are in an n-space; their projections on to the common-factor space are crowded into an r space, and are naturally at smaller angles with one another than the actual test vectors are. These angles between the *projected* test vectors do not, therefore, represent by their cosines the correlations between the tests. The angles are too small for that, and the cosines, therefore, too large. But if we multiply the cosine of such an angle by the lengths of the two projections which it lies between, we again arrive at the correlation.

Thus in Figure 17, the angle between the lines 1 and 3 on the wall is less than the angle between the actual test vectors 1 and 3 out in the 6-space, of which the lines on the wall are the projections. But the lengths of the lines 1 and 3 on the wall are less than the unit length we marked off on the actual vectors, being in fact the roots of the communalities. If we call these lengths on the wall h_1 and h_3, then the product h_1h_3 times the cosine of the projected angle again gives the correlation coefficient.

10. *Rotations.*—It will be remembered that Thurstone, after obtaining a set of loadings for the common factors by his method of analysis of the matrix of correlations, "rotates" the axes until the loadings are all positive—and he also likes to make as many of them as possible zero. It is instructive to look at this procedure in the light of our geometrical picture from which the phrase "rotating the factors" is taken. It should be emphasized first of all that such rotation of the common-factor axes in Thurstone's system must take place entirely within the common-factor space, and the common-factor axes must not leave that space and encroach upon the specifics. In

Figure 16, therefore, no rotation, in Thurstone's sense, of the g axis can be made (since the common-factor space is a line), except indeed reversing its direction and measuring stupidity instead of intelligence.

In Figure 17 the common-factor space is a plane, and the axes a and b can be rotated in this plane, like the hands of a clock fixed permanently at right angles to one another. When the positive directions of a and b enclose all the vector projections, as they do in our figure, then all the loadings are positive. The position shown would, therefore, fulfil this desire of Thurstone's. Moreover, one of the loadings could be made zero, by rotating a and b until a coincides with line 1 (when b will have no loading in Test 1), or until b coincides with line 4 (when a will have no loading in Test 4).

When there are three common factors, the common-factor space is an ordinary 3-space. The three common-factor axes divide this space into eight octants. Rotating them until all the loadings are positive means until all the projections of the test vectors are within the positive octant. This will always be nearly possible if the correlations are all positive. Moreover, it is clear that we can always make at any rate some loadings zero. In the common-factor 3-space we can move one of the axes until it is at right angles to two of the test projections, in which tests that factor will then have no loading. Keeping that axis fixed, we can then rotate the other two axes round it, seeking for a position where one of them is at right angles to some test. The number of zero loadings obtainable will clearly be limited unless the configuration of the test vectors happens to lend itself to many zeros. We shall see later that Thurstone seeks for teams of tests which do this.

Although Thurstone makes his rotations exclusively within the common-factor space, keeping the specifics sacrosanct at their maximum variance, there is, of course, nothing to prevent anyone who does not hold his views from rotating the common-factor axes into a wider space, and increasing the number of common-factor axes at the expense of the specific variance, until ultimately we reach as many common factors as we have tests, and no specifics.

CHAPTER V

HOTELLING'S "PRINCIPAL COMPONENTS"

1. *Another geometrical picture.*—The geometrical picture of the last chapter, however, is not the only form of spatial analogy which can be used for representing the results of mental tests, nor indeed was it the first in the field, though it is the most powerful. The earlier, and perhaps more natural, plan of representing two tests was by two lines at right angles, instead of at an angle depending on their correlation as in Chapter IV. Using the two lines at right angles, and the two test scores as co-ordinates, each person could, in this form of diagram also, be represented by a point on the paper, and his two scores by the feet of perpendiculars from that point on to the test axes. But if, on such a diagram, we mark the points of ten thousand persons, these will, of course, not be distributed in the same circular symmetrical fashion as in Figure 14 (page 55). If we look at Figure 14, we can see what would happen to the crowd of persons if we were to pull the test vectors farther and farther apart * until finally they were at right angles. The shaded northern sector of the crowd is composed of persons whose scores are above average in both tests, and this sector is bounded by the two dotted lines which are at right angles to the test vectors. As the angle between the test vectors grows larger, the two dotted lines in question close towards one another, and this shaded section of the crowd is driven northward. Simultaneously the other shaded section is driven southward. When the test vectors reach a position at right angles to one another, the dotted line at right angles to X falls along Y, and the other along X, and we have Figure 18. The crowd is no longer distributed in a circular fashion round the origin.

* It is understood that they continue to stand for the same tests, with the same correlation, though the latter is no longer represented by the cosine of the angle between the vectors.

It now bulges out to the north and south, in the quadrants where the two test scores are either both positive or both negative, and its lines of equal density, formerly circles, have become ellipses. In this form of diagram, it is this ellipticity of the crowd which shows the presence of correlation between the tests. If the tests are highly correlated, the ellipses will be long and narrow; if they are less correlated, they will be plumper; if there is no correlation, they will be circles; if there is negative correlation, they will be longer the other way, i.e. from east to west in our diagram. In our former figures in Chapter IV, the space of the diagram, whether plane, solid, or multi-dimensional, was peopled by a "spherical" crowd whose density fell away equally in all directions from the origin, while correlation between tests was indicated by the angles between their test vectors. In the figures of the present chapter, all the test vectors are at right angles, and the space is peopled by a crowd whose density falls off differently in different directions unless there is no correlation present.

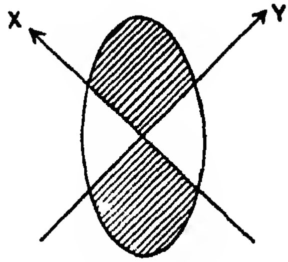

Figure 18.

If we add a third test to the two in Figure 18, its axis, in the present system, has to be at right angles to the first two. The former spherical swarm of persons (of Chapter IV) has become now an ellipsoidal swarm, like a Zeppelin, with proportions determined by the correlations. If these are positive, its greatest length will be in the direction of the positive octant of space (that octant in which all scores are above average, i.e. positive), and the opposite negative octant. Its waist-line will not, as a rule, be circular, but elliptical.

The ellipse of Figure 18 has two principal axes, a major axis from north to south, and a minor axis at right angles to it from east to west. The ellipsoid of three tests has three principal axes; the "ellipsoid" (for we continue to use the term) for n tests will be in n-dimensonal space and will have n principal axes. It is these principal axes of

the ellipsoids of equal density which are the "principal components" of Hotelling's method (Hotelling, 1933). They are exactly equal in number to the tests, but usually the smaller ones are so small as to be negligible, within the limits of exactitude reached by psychological experiment.

2. *The principal axes.*—Finding Hotelling's principal components, therefore, consists in finding those axes, all at right angles to one another, which lie one along each principal axis of the ellipsoids of equal density of the population of persons tested. In Figure 18, for example, one of them lies north and south, the other east and west. The crowd of persons can then be described in terms of these new axes, in terms of factors, that is, instead of in terms of the original tests. These factors are uncorrelated, for the crowd is symmetrically distributed with regard to them, though not in a circular manner. This brings us to one more thing that has to be done to these factors before they become Hotelling's principal components: they have to be measured in new units. The original test scores were, we have tacitly assumed in making our diagrams, measured in comparable units, namely each in units of its own standard deviation. But the factors arrived at by a mere rotation to the principal axes, in an elliptically distributed crowd, are no longer such that the standard deviation of each is represented by the same distance in the diagram. If in Figure 18 all the points representing people are projected on to a horizontal east-and-west factor (Factor II), the feet of these perpendiculars are obviously more crowded together than the corresponding points would be on a north-and-south factor (Factor I). On this diagram, therefore, the standard deviation of Factor II is represented by a shorter distance than is the standard deviation of Factor I. To make these equal, we would have to stretch our paper from east to west, or compress it from north to south, until the crowd was again circular, during which procedure the test vectors would have to move back to the position of Figure 14 to keep the crowd's test scores equal to their projections, and we are then back at the space of Chapter IV. The "ellipsoidal" space of this present chapter, in fact, is used only until the principal

axes of the ellipsoid are discovered, after which, by a change of units along each principal axis, it is made into a "spherical" space again.

In the preceding paragraph, the reader may feel a difficulty which has been known to trouble students in class. If, he may say, we stretch Figure 18 from east to west till the ellipse is a circle, that ought to separate the arrows of the test vectors still farther. Yet you say they will return to the positions shown in Figure 14!

The mistake lies in thinking that stretching the space—the plane of the paper in Figure 18—till the ellipsoid is spherical will move the test vectors with the space. The points representing persons move with the space; indeed, they *are* the space. But the test vectors are not rigidly attached to the space. Each test vector must be such that every person's point, projected on to it, gives his score. If the points move about, as they do when we stretch the paper, the test vector must move so that this remains true, and in our case that means moving nearer together as the crowd becomes more circular. It is just the reverse of the process by which we obtained Figure 18 from Figure 14.

3. *Advantages and disadvantages.*—The advantage of Hotelling's factors can be best appreciated while the crowd of persons is in the ellipsoidal condition. Hotelling's first factor (or "component," as he calls it) runs along the greatest length of the crowd, and gives the best single description of a person's position. If we know all his factor scores, we know exactly where he is in the crowd. If we have to search for him, we would rather be told his position on the long axis, and search along the short ones, than be told his position on any other axis instead. If there are, say, twenty tests, there will be twenty principal axes ranging from longest to shortest, and twenty Hotelling components.* But the first four or five of these will go a long way towards defining a man's position in the tests,

* All that is here said about principal components refers to the case, which is that considered by Hotelling, in which the method of calculation about to be described is applied to the matrix of correlations with unities (or possibly with reliabilities) in the diagonal cells. The method, as a means of calculation, however, could be

and will do so better than any other equally numerous set of factors, whether of Hotelling's or of any other system. In this respect Hotelling's factors undoubtedly stand foremost. They will not, however, reproduce the *correlations* exactly unless they are all used, whereas in Thurstone's system a few common factors can, theoretically, do this, though in actual practice the difference of the two systems in this respect is not great. The chief disadvantage of Hotelling's components is that they change when a new test is added to the battery. When a new test is added to a Spearman battery, *provided that it conforms to the hierarchy*, *g* does not change in nature, though its exactness of measurement is changed. Whether Thurstone's common factors will remain invariant in augmented batteries, and if so under what conditions, is a question we shall consider at a later stage in this book. Though such invariance seems unlikely, it is not obviously inconceivable.

4. *A calculation.*—The actual calculation of the loadings of Hotelling's components requires, for its complete understanding, a grasp of the method of finding algebraically the principal axes of an ellipsoid, a problem which will be found dealt with in three dimensions in any textbook on solid geometry. We give an account of this, for n dimensions, in the Appendix. Here we shall only explain Hotelling's ingenious iterative method of doing this arithmetically, by means of an example, for which we shall use the matrix of correlations already employed in Chapter II to illustrate Thurstone's method (see opposite page).

We have inserted unities in the diagonal cells, for Hotelling's procedure does not contemplate the assumption of specific factors (much less maximized specifics) except possibly that part of a specific which is due to error, in which case what are called " reliabilities " (actual correlations of two administrations of the test) would be used in the diagonal.

Hotelling's arithmetical process then begins with a guess

used to obtain loadings for the common factors after Thurstone's communalities have been inserted, instead of the "centroid" method. But as the factors in the common-factor space have afterwards to be rotated, there would be no point in this use of Hotelling's method.

1·0	·4	·4	·2	·8	·78	·775
·4	1·0	·7	·3	1·0	1·00	1·000
·4	·7	1·0	·3	1·0	1·00	1·000
·2	·3	·3	1·0	·7	·65	·637
·80	·32	·32	·16			
·40	1·00	·70	·30			
·40	·70	1·00	·30			
·14	·21	·21	·70			
1·74	2·23	2·23	1·46			
·780	·312	·312	·156			
·400	1·000	·700	·300			
·400	·700	1·000	·300			
·130	·195	·195	·650			
1·710	2·207	2·207	1·406			

at the proportionate loadings of the first principal component. Practically any guess will do—a bad guess will only make the arithmetic longer. We have guessed ·8, 1, 1, ·7, the numbers to be seen on the right of the matrix, because these numbers are roughly proportional to the sums of the four columns, and such numbers usually give a good first guess.

Each row of the matrix is then multiplied by the guessed number on its right, giving the matrix below the first one, beginning with ·80. We then take, as our second guess, numbers proportional to the sums of the columns of *this* matrix, namely—

	1·74	2·23	2·23	1·46
giving	·78	1	1	·65

That is, we divide the sums of the columns by their largest member, and use the results as new multipliers. They are seen placed farther on the right of the original matrix. It is unusual for two of them to be of the same size—that is a peculiarity of our example.

It is always the original matrix whose rows are multiplied by each improved set of multipliers. The above set gives the next matrix shown, that beginning with ·780, and the sums of *its* columns—

1·710 2·207 2·207 1·406

give a third guess at the multipliers, namely—

·775 1 1 ·687

And so the reiteration goes on, and the reader, who is advised to carry it a stage farther at least, would find if he persevered that the multipliers would change less and less. If he went on long enough, he would reach this point (usually, however, far fewer decimals are sufficient):

	1·0	·4	·4	·2	·772865
	·4	1·0	·7	·3	1·000000
	·4	·7	1·0	·3	1·000000
	·2	·3	·3	1·0	·629811
	·772865	·309146	·309146	·154573	
	·400000	1·000000	·700000	·300000	
	·400000	·700000	1·000000	·300000	
	·125962	·188943	·188943	·629811	
	1·698827	2·198089	2·198089	1·384384	
giving	·772865	1	1	·629813	

that is, totals in exactly the same proportion as the multipliers. These final multipliers (or earlier ones if the experimenter is content with less exact values) are then proportionate to the loadings of the first Hotelling component in the four tests. They have, however, to be reduced until the sum of their squares equals the largest total, 2·198089, which is called the first "latent root" of the original matrix. This is done by dividing them by the square root of the sum of their squares and multiplying them by the square root of the latent root. They then become—

·662 ·857 ·857 ·540.

The next step in Hotelling's process is similar to one with which we have already become familiar in Thurstone's method. The parts of the variances and correlations due to this first component are calculated and subtracted from the original experimental matrix. These variances and correlations due to the first component are:

	·662	·857	·857	·540		
·662	·439	·567	·567	·357		
·857	·567	·734	·734	·462	Matrix due to	
·857	·567	·734	·734	·462	first principal	
·540	·357	·462	·462	·291	component.	
	·561	− ·167	− ·167	− ·157	·3	·18
Residual	− ·167	·266	− ·034	− ·162	− ·4	− ·38
matrix	− ·167	− ·034	·266	− ·162	− ·4	− ·38
	− ·157	− ·162	− ·162	·709	1·0	1·00
	·168	− ·050	− ·050	− ·047		
	·067	− ·106	·013	·065		
	·067	·013	− ·106	·065		
	− ·157	− ·162	− ·162	·709		
	·145	− ·305	− ·305	·792		

The residual matrix is then treated in exactly the same way as the original matrix, the beginnings of the process being shown above. The guessed multipliers, proportional to the sums of the columns, are not so near the truth this time, for the first one, which we have guessed at ·3, and which reduces after one operation to ·18, goes on reducing until it becomes negative, the final values of these second loadings being as shown in the appropriate column of the following table, which also gives the loadings of the third and fourth factors, obtained in the same way. The variances and correlations due to each factor in turn are subtracted from the preceding residual matrix and the new residual matrix analysed for the next factor :

Factor	*I*	*II*	*III*	*IV*	*Sum of Squares*
Test 1	·662218	− ·323324	·675967	.	1
,, 2	·856836	− ·135197	− ·312332	− ·387298	1
,, 3	·856836	− ·135197	− ·312332	·387298	1
,, 4	·539645	·826092	·162323	.	1
Sum of squares *	2·198090	·823526	·678383	·300000	4
Percentages	55·0	20·6	16·9	7·5	100

* These four quantities are, in the Hotelling process, what are called the "latent roots" of the matrix.

An alternative method of finding principal components, due to Kelley, is to deal with the variables two at a time. The pair first chosen are rotated in their plane until they are uncorrelated. Then the same is done to another pair, and so on, the new uncorrelated variables being in turn paired with others, until finally all correlations are zero. (Kelley, 1935, Chapters I and VI.) A chief advantage is that the components are obtained *pari passu*, and not successively; also, in certain circumstances where Hotelling's process converges very slowly, Kelley's is quicker. The end results are the same.

5. *Acceleration by powering the matrix.*—In a later paper Hotelling pointed out that his process of finding the loadings of the principal components can be much expedited by analysing, not the matrix of correlations itself, but its square, or fourth, eighth, or sixteenth power, got by repeated squaring (Hotelling, 1935*b*). Squaring a symmetrical matrix is a special case of matrix multiplication (see Chapter VII, Section 8): it is done by finding the "inner products" (see footnote, page 31) of each pair of rows, including each row with itself, and setting the results down in order. Applying this to the correlation matrix:

$$\begin{bmatrix} 1{\cdot}0 & {\cdot}4 & {\cdot}4 & {\cdot}2 \\ {\cdot}4 & 1{\cdot}0 & {\cdot}7 & {\cdot}3 \\ {\cdot}4 & {\cdot}7 & 1{\cdot}0 & {\cdot}3 \\ {\cdot}2 & {\cdot}3 & {\cdot}3 & 1{\cdot}0 \end{bmatrix}$$

we see that the inner product of the first row with itself is 1·36; of the first row with the second, 1·14; and so on. Setting these down in order, we get for the matrix squared:

$$\begin{bmatrix} 1{\cdot}36 & 1{\cdot}14 & 1{\cdot}14 & {\cdot}64 \\ 1{\cdot}14 & 1{\cdot}74 & 1{\cdot}65 & {\cdot}89 \\ 1{\cdot}14 & 1{\cdot}65 & 1{\cdot}74 & {\cdot}89 \\ {\cdot}64 & {\cdot}89 & {\cdot}89 & 1{\cdot}22 \end{bmatrix}$$

Exactly the same process is applied to this, beginning with guessed multipliers, as we applied to the original matrix. The multipliers, however, settle down twice as rapidly towards their final values, *which are the same here as there.* We have finally:

	1·36	1·14	1·14	·64	·772865
	1·14	1·74	1·65	·89	1·000000
	1·14	1·65	1·74	·89	1·000000
	·64	·89	·89	1·22	·629811
	1·051096	·881066	·881066	·494634	
	1·140000	1·740000	1·650000	·890000	
	1·140000	1·650000	1·740000	·890000	
	·403079	·560532	·560532	·768369	
	3·784175	4·831598	4·831598	3·043003	
Ratio	·772865	1·000000	1·000000	·629812	

The "latent root," however, or largest total, 4·831598, is the square of the former latent root, 2·198090, so that its square root must be taken before we complete finding the loadings.

In exactly the same way the squared matrix may be again squared, and again and again, before we analyse it. The more we square it, the quicker the Hotelling iteration process works. The end multipliers are always the same, but the " root " is the same power of the root we need as is the matrix of the original matrix.

A still further acceleration of the process is due to Cyril Burt, who observed that as the matrix is repeatedly squared it becomes more and more nearly hierarchical, including the diagonal cells (Burt, 1937*a*). This is due to the largest factor increasingly predominating as it is " powered," especially if the largest latent root is widely separated from the others. In consequence, the square roots of the diagonal cells become more and more nearly in the ratio of the Hotelling multipliers, and form an excellent first guess for the latter. When our matrix is squared twice again, giving the eighth power, it becomes:

$$\begin{bmatrix} 108\cdot78 & 140\cdot67 & 140\cdot67 & 88\cdot54 \\ 140\cdot67 & 182\cdot03 & 182\cdot03 & 114\cdot61 \\ 140\cdot67 & 182\cdot03 & 182\cdot03 & 114\cdot61 \\ 88\cdot54 & 114\cdot61 & 114\cdot61 & 72\cdot38 \end{bmatrix}$$

and the square roots of its diagonal members are—

10·429	13·492	13·492	8·508

which are in the ratio—

·7730	1	1	·6306

very near indeed to the Hotelling final multipliers—

·772865	1	1	·629811

Hotelling gives a method of finding the residues, for the purpose of calculating the next factor loadings, from the "powered" matrix. But it may be so nearly perfectly hierarchical that this fails unless an enormous number of decimals have been retained, and it is in practice best to go back to the original matrix and obtain the residues from it. Their matrix can in turn be squared, and so on. Other and very powerful methods of acceleration will be found described in Aitken, 1937*b*.

6. *Properties of the loadings.*—If all the Hotelling components are calculated accurately, their loadings ought completely to exhaust the variance of each test; that is, the sum of the squares of the loadings in each row should be unity. The sum of the squares of the loadings in each column equals the "latent root" corresponding to that column, and the sum of the four latent roots is exactly equal to the number of tests. Each latent root represents the part of the whole variance of all the tests which has been "taken out" by that factor. Thus the first factor "takes out" 55 per cent., the first two factors together 75·6 per cent., of the variance of the original scores. The four factors account for all the variance.

If we turn back to Chapter II, where we made a Thurstone analysis of this same battery of four tests into two common factors and four specifics (six factors in all), we see, in the table on page 35, that the two first Thurstone factors

"take out" 1·7652 and ·0515 respectively—that is, 44·1 per cent. and 1·3 per cent. of the four tests—much less than the two first Hotelling factors account for. Because of this, the two first Hotelling factors will reproduce the original scores much better than the two Thurstone factors will. On the other hand, the two Thurstone factors reproduce the correlations exactly, while it takes all four Hotelling factors to do this.

The correlations which correspond to the loadings given in the table on page 78 are obtained by finding the "inner product" of each pair of rows. Applying this to the table we find the correlation r_{24}, say, to be—

$$\cdot 856836 \times \cdot 539645 - \cdot 135197 \times \cdot 826092 - \cdot 312332 \times \cdot 162823 - \cdot 387298 \times \text{zero} = \cdot 300000$$

In this way, as we said above, the loadings of the four Hotelling factors will exactly reproduce the correlations we began with. If, however, we have stopped the analysis after we have found only two principal components (or factors), these two would have reproduced the correlations only approximately. For example, for r_{24} we should only have—

$$\cdot 856836 \times \cdot 539645 - \cdot 135197 \times \cdot 826092 = \cdot 350702 \text{ instead of } \cdot 300000$$

Before we leave the table of Hotelling loadings, we may note that the signs of any column of the loadings can be reversed without changing either the variances or the correlations. Reversing the signs in a column merely means that we measure that factor from the opposite end, as we might rank people either for intelligence or stupidity and get the same order, but reversed. We will usually desire to call that direction of a factor positive which most conforms with the positive direction of the tests themselves, and therefore we will usually make the largest loading in each column positive.

All the loadings of Hotelling's first factor are, in an ordinary set of tests, positive. Of the other loadings, about half are negative. Thurstone's first analysis, it will be remembered, also gave a number of negative loadings

to the factors after the first, but he rotated his factors until these disappeared. That cannot be done here, or the principal components would lose their virtue of being the principal axes of the ellipsoids of density.

7. *Calculation of a man's principal components. Estimation unnecessary.*—The Hotelling components have one other advantage over other kinds of factors that we did not mention in Section 3. *They can be calculated exactly* from a man's scores, whereas Spearman or Thurstone factors can only be estimated. This is because the Hotelling components are never more numerous than the tests, whereas the Thurstone or Spearman factors, including the specifics, are always more numerous than the tests. For the Hotelling components, therefore, we always have just the same number of equations as unknowns, whereas we have more unknowns than equations in the Spearman-Thurstone system.

We have hitherto given the analysis of tests into factors in the form of tables of loadings, or *matrices* of loadings, as we may call them, adopting the mathematical term. But we can alternatively write them out as "specification equations," as we shall call them. Thus the table on page 73 would be written—

$$\begin{aligned}
z_1 &= \cdot 662218\gamma_1 - \cdot 828824\gamma_2 + \cdot 675967\gamma_3 \\
z_2 &= \cdot 856836\gamma_1 - \cdot 135197\gamma_2 - \cdot 312882\gamma_3 - \cdot 387298\gamma_4 \\
z_3 &= \cdot 856836\gamma_1 - \cdot 135197\gamma_2 - \cdot 312882\gamma_3 + \cdot 387298\gamma_4 \\
z_4 &= \cdot 539645\gamma_1 + \cdot 826092\gamma_2 + \cdot 162323\gamma_3
\end{aligned}$$

Here z_1, z_2, z_3, and z_4 stand for the scores in the four tests, measured in standard units; that is, measured from the mean in units of standard deviation. The factors γ_1, γ_2, γ_3, and γ_4 are also supposed to be measured in such units. These specification equations enable us to calculate any man's standard score in each test if we know his factors, and since there are just as many equations as factors, they can be solved for the γ's and enable us to calculate, conversely, any man's factors if we know his scores in the tests. The solution to these Hotelling equations for the γ's happens to be peculiarly simple, as we shall prove in the Appendix, Section 7. It is as follows—

$$
\begin{aligned}
\gamma_1 &= (\;\;\cdot662218z_1 + \cdot856886z_2 + \cdot856886z_3 + \cdot539645z_4) \div 2\cdot198090\\
\gamma_2 &= (-\cdot323324z_1 - \cdot135197z_2 - \cdot135197z_3 + \cdot826092z_4) \div \cdot823526\\
\gamma_3 &= (\;\;\cdot675967z_1 - \cdot312382z_2 - \cdot312382z_3 + \cdot162328z_4) \div \cdot678383\\
\gamma_4 &= (\qquad\qquad\quad - \cdot387298z_2 + \cdot387298z_3 \qquad\qquad\quad) \div \cdot300000
\end{aligned}
$$

The table on page 78, therefore, serves a double purpose. Read horizontally it gives the composition of each test in terms of factors. Read vertically it gives the composition of each factor in terms of tests, if we divide the result by the root at the foot of the column.*

Suppose, for example, that a man or child has the following scores in the four tests—

$$1\cdot29 \qquad \cdot36 \qquad \cdot72 \qquad 1\cdot03.$$

This is evidently a person above the average in each test, since the scores are all positive. His factors will be obtained by substituting these scores for the z's in the above equations, with the result—

$$
\begin{aligned}
\gamma_1 &= 1\cdot062504\\
\gamma_2 &= \;\;\cdot349441\\
\gamma_3 &= 1\cdot034624\\
\gamma_4 &= \;\;\cdot464757
\end{aligned}
$$

(Of course, in practical work six decimal places would be absurd. They are given here because we are using this artificial example to illustrate theoretical points, in place of doing algebraic transformations, and they need, therefore, to be exact.)

If these values for the factors are now inserted in the specification equations opposite, the scores z in the test will be reproduced exactly ($1\cdot29$, $\cdot36$, $\cdot72$, and $1\cdot03$).

Notice, too, that if we have stopped our analysis at less than the full number of Hotelling factors, we can nevertheless calculate these factors for any person exactly. As soon as we have the first column of the table on page 78, we can calculate γ_1 for anyone whose scores z we know.

Had we done this with the person whose scores are given

* If the analysis has been performed with "reliabilities" in the diagonal cells instead of units, the statement in the text still holds (Hotelling, 1933, 498). If on correlations corrected for "attenuation," the matter is more complicated (*ibid.* 499–502).

above, we should have summarized his ability in these four tests by the one statement—

$$\gamma_1 = 1{\cdot}062504$$

This would have been an incomplete statement, but it is the best single statement that can be arrived at. If we attempt to reproduce the scores z from this one factor alone, we can use only the first term in each of the specification equations on page 78. These give for the scores—

	·704	·910	·910	·573	
instead of	1·29	·36	·72	1·03,	the true values,

a pretty bad shot, as the reader will agree. But bad as it is, it is better than any other one factor will provide, as we shall show later after we have considered how to estimate Spearman and Thurstone factors.

It will be seen from these first chapters that the different systems of factors proposed by different schools of "factorists" have each their own advantages and disadvantages, and it is really impossible to decide between them without first deciding why we want to make factorial analyses at all. This fundamental question we will devote some pages to in later chapters. But there are still several things we must do in preparation, and we turn next to a matter which has wider applications than in factorial analysis, namely the method of estimating one quality from measurements of other qualities with which it is correlated. This, for example, is the problem before those who give vocational advice to a man after putting him through various tests, or who give educational advice (or more peremptory instructions) to English children of eleven years of age after examining them in English, arithmetic, and perhaps with an "intelligence test," sorting them into those who may attend a secondary school, those who go to a central school, and those who remain in an elementary school.

PART II

THE ESTIMATION OF FACTORS

To simplify and clarify the exposition, errors due to sampling the population of persons are in Parts I and II assumed to be non-existent.

CHAPTER VI

ESTIMATION AND THE POOLING SQUARE

1. *Correlation coefficient as estimation coefficient.*—A correlation coefficient indicates the degree of resemblance between two lists of marks: and therefore it also indicates the confidence with which we can estimate a man's position in one such list x if we know his position in the other y. If the correlation between two lists is perfect ($r = 1$), we know that his standardized score * in the one list is exactly the same as in the other ($x = y$).

If the correlation between the two lists is zero ($r = 0$), then the knowledge of a man's position in the one list tells us nothing whatever about his position in the other list. If we are *compelled* to make an estimate of that, we can only fall back on our knowledge that most men are near the average and few men are very good or very bad in any quality. We have, therefore, most chance of being correct if we guess that this man is average in the unknown test. ($x = 0$. The average mark we have agreed to call zero; marks above average, positive; marks below average, negative.)

In the first case, when $r = 1$, we are justified in equating his unknown score x to his known score y—

$$x = y$$

In the second case, when $r = 0$, we are compelled by our ignorance to take refuge in—

$$x = 0 \text{ or average.}$$

Both these statements can be summed up in the one statement—

$$\hat{x} = ry$$

where the circumflex mark over the x is meant to indicate

* A test score always means a standardized score unless the contrary is stated. But *estimates* are not in standard measure in general.

that this is an estimated, not a measured, value. If, now, we consider a case between these, where the correlation is neither perfect nor zero, it can be shown that this equation still holds, provided each score is measured in standard deviation units. Since r is always a fraction, this means that we always estimate his unknown x score as being nearer the average than his known y score. That is because we know that *men tend to be average men*. If this man's y score is high, say—

$$y = 2$$

(two standard deviations above the average), and if the correlation between the qualities x and y is known to be $r = \cdot 5$, we guess his position in the x test as being—

$$\hat{x} = ry = \cdot 5 \times 2 = 1$$

i.e. only one standard deviation above the average. This is a guess influenced by our two pieces of knowledge, (1) that he did very well in Test y, which is correlated with Test x, and (2) that most men get round about an average score (zero). It is a compromise, an estimate. It will often be wrong; indeed, very seldom will it be exactly right. But it will be right *on the average*, it will as often be an underestimate as an overestimate, in each array of men who are alike in y. The correlation coefficient, then, is an estimation coefficient for tests measured in standard deviation units.

2. *Three tests.*—Suppose now that we have *three* tests whose intercorrelations are known, and that a man's scores on two of them, y and z, are known. We wish to estimate what his score will most probably be in the other test, x. x need not be a test in the ordinary sense of the word, but may be an occupation for which the man is a candidate or entrant. According as we use his known y or his known z score, we shall have two estimates for his x score. To fix our ideas, let us take definite values for the correlations, say:

	x	y	z
x	1·0	·7	·5
y	·7	1·0	·8
z	·5	·8	1·0

The two estimates for his x are then—

$$\hat{x} = \cdot 7y$$
$$\hat{x} = \cdot 5z$$

and of these we shall have rather more confidence in the estimate associated with the higher correlation. But we ought to have still more confidence in an estimate derived from both y and z. Such an estimate could use not only the knowledge that y and z are correlated with x, but also the knowledge that they are correlated to an extent of $r = \cdot 3$ with each other. Just to take the average of the above two separate estimates will not utilize this knowledge, nor will it utilize the fact that the estimate from y ($r = \cdot 7$) is more worthy of confidence than the estimate from z ($r = \cdot 5$).

What we want is to know how to combine the two scores y and z into a weighted total—

$$(by + cz)$$

which will have the highest possible correlation with x. Such a correlation of a *best*-weighted total with another test is called a *multiple* correlation. From such a weighted total of his two known scores we could then estimate the man's x score more accurately than from either the y or the z score alone. It must use all the information we have, including our information that y and z correlate to an amount $r = \cdot 3$.

3. *The straight sum and the pooling square.*—In order to answer this question, we shall first consider the problem of finding the correlation of the straight unweighted sum of the scores $y + z$ with x. This is the simplest form of a problem to which a general answer was given by Professor Spearman (Spearman, 1913).

We shall put his formula into a very simple form, which we may call a *pooling square.* In our present instance we want to find the correlation of $y + z$ with x (all of these being, we are assuming, measured in standard deviation units). We divide the matrix of correlations by lines separating the " criterion " x from the " battery " $y + z$ thus:

	x	y	z
x	1·0	·7	·5
y	·7	1·0	·3
z	·5	·3	1·0

In each of the quadrants of this pooling square (with unities in the diagonal, be it noted) we are going to form the sum of all the numbers, and we shall indicate these sums by the letters :

$$\begin{array}{c|c} A & C \\ C & B \end{array}$$

(where C is the sum of the Cross-correlations between the battery $y + z$ and the criterion x, which can be regarded as a second battery of one test only).

Then the correlation of x with $y + z$ is equal to—

$$\frac{C}{\sqrt{AB}}$$

which in our present example is—

$$\frac{\cdot 7 + \cdot 5}{\sqrt{(1) \times (1 + \cdot 3 + \cdot 3 + 1)}} = \frac{1 \cdot 2}{\sqrt{2 \cdot 6}} = \cdot 744$$

so that the battery $(y + z)$ has a rather better correlation (·744) with x than has either of its members (·7 and ·5). From the straight sum of the man's scores in the two tests y and z we can therefore in this case get a better estimate of his score in x than we could get from either alone.

4. *The pooling square with weights.*—We want, however, to know whether a *weighted* sum of y and z will give a still higher combined correlation with x. With sufficient patience, we could answer this by trial and error, for the pooling square enables us to find almost as easily the correlation of a weighted battery with the criterion.* Let us, for example, try the battery $3y + z$. For this purpose

* The pooling square can also be used to find the correlations or covariances of weighted batteries with one another. Elegant developments are Hotelling's ideas of the most predictable criterion (1935*a*) and of vector correlation (1936).

we write the weights along both margins of the pooling square :

$$\begin{array}{cc|cc}
 & & 3 & 1 \\
\hline
 & 1{\cdot}0 & {\cdot}7 & {\cdot}5 \\
\hline
3 & {\cdot}7 & 1{\cdot}0 & {\cdot}3 \\
1 & {\cdot}5 & {\cdot}3 & 1{\cdot}0
\end{array}$$

and multiply *both the rows and the columns* by these weights before forming the sums A, B, and C. The result of the multiplications in our case is :

$$\begin{array}{c|cc}
1{\cdot}0 & 2{\cdot}1 & {\cdot}5 \\
\hline
2{\cdot}1 & 9{\cdot}0 & {\cdot}9 \\
{\cdot}5 & {\cdot}9 & 1{\cdot}0
\end{array}
\quad = \quad
\begin{array}{c|c}
1{\cdot}0 & 2{\cdot}6 \\
\hline
2{\cdot}6 & 11{\cdot}8
\end{array}$$

and we therefore have—

$$\text{correlation} = \frac{2{\cdot}6}{\sqrt{11{\cdot}8}} = {\cdot}757$$

a higher value than ·744 given by the simple sum. So we have improved our estimation of the man's x score, and estimates made by taking $3y + z$ would correlate ·757 with the measured values of x.

5. ***Regression coefficients and multiple correlation.***—Similarly we could try other weights for y and z and search by trial and error for the best. There is, however, a general answer to this question, namely that *the best weights for y and z are proportional to certain minor determinants of the correlation matrix.* The weight for y is proportional to the minor left when we cross out the criterion column and the y row, the weight for z is proportional to *minus* the minor left when we similarly cross out the criterion column and the z row. The matrix of correlations with the criterion column deleted being :

$$\begin{bmatrix}
{\cdot}7 & {\cdot}5 \\
1{\cdot}0 & {\cdot}3 \\
{\cdot}3 & 1{\cdot}0
\end{bmatrix}$$

the weight for y is therefore proportional to:

$$\begin{vmatrix} \cdot 7 & \cdot 5 \\ \cdot 3 & 1\cdot 0 \end{vmatrix} = \cdot 55$$

and that for z is proportional to:

$$-\begin{vmatrix} \cdot 7 & \cdot 5 \\ 1\cdot 0 & \cdot 3 \end{vmatrix} = \cdot 29$$

that is, they are as ·55 : ·29. To make these weights not merely proportional but absolute values we must divide each of them by the minor left when the row and column concerned with the "criterion" x are deleted, namely:

$$\begin{vmatrix} 1\cdot 0 & \cdot 3 \\ \cdot 3 & 1\cdot 0 \end{vmatrix} = \cdot 91$$

so that these absolute best weights, for which the technical name is "regression coefficients," are—

$$\frac{\cdot 55}{\cdot 91}y + \frac{\cdot 29}{\cdot 91}z$$

or $$\cdot 6044y + \cdot 3187z$$

We are inviting the reader to take this method of calculating the regression coefficients on trust; but he can at least satisfy himself that when applied to the pooling square they give a higher correlation of battery with criterion than any other weights do. The result of multiplying the y column and row by ·6044, and the z column and row by ·3187, is the following:

$$\begin{array}{c} \\ \cdot 6044 \\ \cdot 3187 \end{array}\begin{array}{ccc} & \cdot 6044 & \cdot 3187 \\ \hline 1\cdot 0 & \cdot 7 & \cdot 5 \\ \cdot 7 & 1\cdot 0 & \cdot 3 \\ \cdot 5 & \cdot 3 & 1\cdot 0 \end{array} = \begin{array}{c|cc} 1\cdot 0000 & \cdot 4231 & \cdot 1593 \\ \hline \cdot 4231 & \cdot 3653 & \cdot 0578 \\ \cdot 1593 & \cdot 0578 & \cdot 1015 \end{array}$$

$$= \begin{array}{c|c} 1\cdot 0000 & \cdot 5824 \\ \hline \cdot 5824 & \cdot 5824 \end{array}$$

Multiple correlation $= \dfrac{\cdot 5824}{\sqrt{\cdot 5824}} = \cdot 763 = r_m$, say, which is higher than any other weighting will produce, if the reader cares to try others. Notice the peculiarity of the pooling

square with regression coefficients as weights, that $C = B$ (·5824 = ·5824). We can deduce that the inner product of the regression coefficients with the correlation coefficients gives the square of the multiple correlation—

$$\cdot 604 \times \cdot 7 + \cdot 319 \times \cdot 5 = \cdot 583 = r_m^2$$

Indeed, we can take this as forming one reason for using ·604 and ·319, and not any other numbers proportional to them, although the latter would give the same order of merit. We want our estimates of x not merely to be as highly correlated with the true values of x as is possible, but also to be equal to them on the average in the long run, in the sense that our overestimations will, in each array of men who have the same y and z, be as numerous as our underestimations, and this is achieved by using not merely ·55 and ·29 as weights, but ·55 ÷ ·91, and ·29 ÷ ·91.

6. ***Aitken's method of pivotal condensation.***—When there are more than two tests y and z in the battery, the application of the above rules becomes increasingly laborious. It is desirable, therefore, to have a routine method of calculating regression coefficients which will give the result as easily as possible even in the case of a team of many tests. The method we shall adopt (Aitken, 1937*a*) is based upon the calculation of tetrads, as already used in our Chapter II. We shall first calculate the above regression coefficients again by this method. Delete the criterion column in the matrix of correlations, *transfer the criterion row to the bottom*, and write the resulting oblong matrix in the top left-hand corner of the sheet of calculations, preferably on paper ruled in squares:

					Check Column
	(1·0)	·3	−1	.	·3
A	·3	1·0	.	−1	·3
	·7	·5	.	.	1·2
B		(·91)	·3	−1	·21
		1·00	·3297	−1·0989	·2308
		·29	·7	.	·99
C			·604	·319	·923

On the right of the oblong matrix of correlation coefficients we rule a middle block of columns of the same number, here *two*, and on the right of all a check column. The columns of the middle block we fill with a pattern of *minus ones* diagonally as shown, leaving the other cells empty,* including the bottom row. In the check column we write the sum of each row. The top left-hand number of all we mark as the "pivot." Slab *B* of the calculation is then formed from slab *A* by writing down, in order as they come, all the tetrad-differences of which the pivot in *A* is one corner. Thus the first row of slab *B* is calculated thus—

$$\begin{aligned}
1 \times 1 - \cdot 3 \times \cdot 3 &= \cdot 91 \\
1 \times 0 - \cdot 3 \times (-1) &= \cdot 3 \\
1 \times (-1) - \cdot 3 \times 0 &= -1 \\
1 \times \cdot 3 - \cdot 3 \times \cdot 3 &= \cdot 21
\end{aligned}$$

and the row is checked by noting that ·21 is the sum of the others. Immediately below this first row a second version of it is written, with every member divided by the first (·91). This is to facilitate the calculation of slab *C* by having unity again as a pivot. The second row of slab *B* is then formed, beginning with—

$$1 \times \cdot 5 - \cdot 7 \times \cdot 3 = \cdot 29$$

Throughout the whole calculation, except for the division of the first row, only one operation needs to be performed, namely the computing of tetrad-differences, beginning with the pivot.

The same operation is then repeated to give slab *C*, using the modified first row of *B*, with pivot unity.

This procedure goes on, slab after slab, until no numbers remain in the left-hand block. There being only three tests in all in our example, this happens at slab *C*. The middle block then gives the regression coefficients ·604 and ·319, with their proper signs, all ready for use. Throughout the calculation the check column detects any blunder in each row.

When the number of tests in the battery is large, the

* The dots represent zeros.

calculation of the regression coefficients is a laborious business, but probably less so by this method than by any other. It will be clear to the reader that so long a calculation is not worth performing unless the accuracy of the original correlation coefficients is high. Only very accurate values can stand such repeated multiplication, etc., without giving untrustworthy results (Etherington, 1932). In other words, regression coefficients have a rather high standard error.

7. *A larger example.*—Next we give in full the calculation of the regression coefficients in a slightly larger example, though one still much smaller than a practical scheme of vocational advice would involve. Here z_0 is the "occupation," and z_1, z_2, z_3, and z_4 are tests. To give the example an air of reality, these and their intercorrelations are taken from Dr. W. P. Alexander's experimental study, *Intelligence, Concrete and Abstract* (Alexander, 1935). They were * :

z_1 Stanford-Binet test ;
z_2 Thorndike reading test ;
z_3 Spearman's analogies test in geometrical figures ;
z_4 A picture-completion test.

But the occupation is a pure invention, for purposes of this illustration only. The correlation matrix is :

	z_0	z_1	z_2	z_3	z_4
z_0	1·00	·72	·58	·41	·63
z_1	·72	1·00	·69	·49	·39
z_2	·58	·69	1·00	·38	·19
z_3	·41	·49	·38	1·00	·27
z_4	·63	·39	·19	·27	1·00

The fact that we possess these correlations means that we have given these tests to a sufficiently large number of

* In this, as in other instances where data for small examples are taken from experimental papers, neither criticism nor comment is in any way intended. Illustrations are restricted to few tests for economy of space and clearness of exposition, but in the experiments from which the data are taken many more tests are employed, and the purpose may be quite different from that of this book.

persons whose ability in the occupation is also known. The occupation can be looked upon as another test, in which marks can be scored. In an actual experiment, obtaining marks for these persons' abilities in the occupation is in fact one of the most difficult parts of the work. We can now find by Aitken's method the best weights for Tests z_1 to z_4 to make their weighted sum correlate as highly as possible with z_0. To make the arithmetic as easy as possible to follow in an illustration, the original correlation coefficients are given to two places of decimals only, and only three places of decimals are kept at each stage of the calculation. The previous explanation ought to enable the reader to follow. As an additional help, take the explanation of the value ·153 in the middle of slab *D*. It is obtained thus from slab *C*—

$$1 \times \cdot 158 - \cdot 050 \times \cdot 106 = \cdot 153$$

and is typical of all the others. Except for the division of each first row, only one kind of operation is required through the whole calculation, which becomes quite mechanical. The numbers shown on the left in brackets are the reciprocals of ·524, ·757, ·826, used as multipliers instead of dividing by the latter numbers, in obtaining the modified first rows. The process continues until the left-hand block is empty, when the regression coefficients appear in the middle block (see opposite page).

The result is that we find that the best prediction of a man's probable success in this occupation is given by the regression equation—

$$\hat{z}_0 = \cdot 390 z_1 + \cdot 222 z_2 + \cdot 018 z_3 + \cdot 481 z_4$$

We give a candidate the four tests, reduce his scores

* The product of all the unconverted pivots, $1 \times \cdot 524 \times \cdot 757 \times \cdot 826$, is the value ·328 of the determinant:

$$\begin{vmatrix} 1\cdot 00 & \cdot 69 & \cdot 49 & \cdot 39 \\ \cdot 69 & 1\cdot 00 & \cdot 38 & \cdot 19 \\ \cdot 49 & \cdot 38 & 1\cdot 00 & \cdot 27 \\ \cdot 39 & \cdot 19 & \cdot 27 & 1\cdot 00 \end{vmatrix}$$

If this alone were wanted, the middle block, and the criterion bottom row, would of course be omitted.

COMPUTATION OF REGRESSION COEFFICIENTS

Aitken's Modified Method with Each Pivot converted to Unity

									Check
A	(1)	·69	·49	·39	−1	.	.	.	1·57
	·69	1	·38	·19	.	−1	.	.	1·26
	·49	·38	1	·27	.	.	−1	.	1·14
	·39	·19	·27	1	.	.	.	−1	·85
	·72	·58	·41	·63	.	.	.	.	2·34
	(1·908)	(·524)	·042	−·079	·690	−1	.	.	·177
B		1·000	·080	−·151	1·317	−1·908	.	.	·338
		·042	·760	·079	·490	.	−1	.	·371
		−·079	·079	·848	·390	.	.	−1	·238
		·083	·057	·349	·720	.	.	.	1·210
	(1·321)		(·757)	·085	·435	·080	−1	.	·357
C			1·000	·112	·575	·106	−1·321	.	·472
			·085	·836	·494	−·151	.	−1	·265
			·050	·362	·611	·158	.	.	1·182
	(1·211)			(·826)	·445	−·160	·112	−1	·225
D				1·000	·539	−·194	·136	−1·211	·272
				·356	·582	·153	·066	.	1·158
E					·390	·222	·018	·431	1·061
					Regression Coefficients				

to standard measure by dividing by the known standard deviation of each test, insert these standard scores into this equation, and obtain an estimated score for him in the occupation. Thus the following three young men could be placed in their probable order of efficiency in this occupation from their test scores:

	Standard Scores in				
	z_1	z_2	z_3	z_4	$\hat{z}_0$
Tom	·7	·2	− ·5	·0	·31
Dick	− ·4	·1	·3	− ·8	− ·47
Harry	·2	·8	·6	1·3	·83

The multiple correlation of such estimates $\hat{z}_0$ with the true values would be obtained by inserting the four correlation coefficients—

$$\cdot 72 \qquad \cdot 58 \qquad \cdot 41 \qquad \cdot 63$$

instead of the z's in the regression equation, and taking the square root, thus—

$$\cdot 390 \times \cdot 72 + \cdot 222 \times \cdot 58 + \cdot 018 \times \cdot 41 + \cdot 431 \times \cdot 63$$
$$= \cdot 68847 = r_m^2$$
$$\therefore r_m = \cdot 83$$

Finally, we can, as we did in the former example, use the regression weights on a pooling square and see if we obtain this same multiple correlation of $r_m = \cdot 83$:

		·390	·222	·018	·431
	1·00	·72	·58	·41	·63
·390	·72	1·00	·69	·49	·39
·222	·58	·69	1·00	·38	·19
·018	·41	·49	·38	1·00	·27
·431	·63	·39	·19	·27	1·00

It will be remembered that we have to multiply each row and column by its appropriate weight, and then sum all the numbers in each quadrant. The easiest way of doing this in large pooling squares is to multiply the rows first, then add the columns and multiply the totals by the column weights, finally adding these products, thus:

Multiply the rows:

		·390	·222	·018	·431
	1·0000	·72	·58	·41	·63
	·2808	·3900	·2691	·1911	·1521
	·1288	·1532	·2220	·0844	·0422
	·0074	·0088	·0068	·0180	·0049
	·2715	·1681	·0819	·1164	·4310
Sums	·6885	·7201	·5798	·4093	·6302

If we had kept all decimals these columnar sums would, since we are using regression coefficients as weights, have been exactly equal to the top row. With the actual figures shown, on multiplying the column totals and adding them, we find that the pooling square condenses to:

1·0000	·6885
·6885	·6885

$$r_m = \frac{\cdot 6885}{\sqrt{\cdot 6885}} = \cdot 83 \text{ as before.}$$

8. *The geometrical picture of estimation.*—Before we close this chapter it will be illuminating to consider what estimation of occupational ability means in terms of the geometrical picture of Chapter IV. Consider the illustration used in the earlier pages of the present chapter, with the matrix:

	x	y	z
x	1·0	·7	·5
y	·7	1·0	·3
z	·5	·3	1·0

Here x is the criterion, y and z are the tests. Each of them can be represented by a line vector, as explained in Chapter IV, with angles between these vectors such that their cosines are the above correlations. The three vectors will then be in an ordinary space of three dimensions.

The *two* tests y and z themselves have, of course, vectors which lie in a plane: any two lines springing from the same point as origin lie in a plane. These are the two tests to which we subject the candidate, whose probable score in x we are then going to estimate. His two scores OY and OZ in y and z enable us to assign to this man a point P on the yz plane, a point so chosen that its projections on to the y and z vectors give the scores made by him in those tests (see Figure 19). ***But we cannot say that this is his point in the three-dimensional space of x, y, and z.*** His point in that space may be anywhere on a line $P'PP''$

at right angles to the plane yz. For from anywhere on that line, projections on to y and z fall on the points Y and Z. Yet the projection on to the vector x, which gives his score in the criterion test x, depends very much on the position of his point on the line $P'PP''$. All the people represented by points on that line have the same scores in y and z but different scores in x, and our man may be any one of them. Before deciding what to do in these

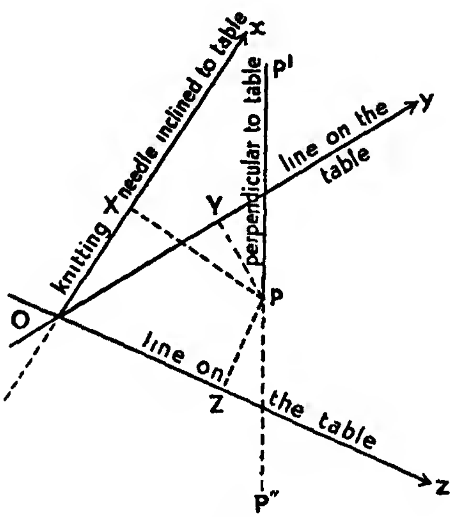

Figure 19.

circumstances, let us consider this set of people $P'PP''$ in more detail.

It will be remembered that the whole population of persons is represented by a spherical swarm of points, crowded together most closely round about the origin O, and falling off in density equally in all directions from that point. Every test vector is a diameter of this sphere, and the plane containing any two test vectors divides the spherical swarm into equal hemispheres. It follows that a line like $P'PP''$ is a chord of the sphere at right angles to a diameter (the line OP), and consequently that it is peopled symmetrically on both sides of P, both upwards along PP' in our figure, and downwards along PP'', the

men on the line being most crowded near the point P itself. The average man of the array of men $P'PP''$ (who are all alike in their scores in the two tests y and z) is therefore the man at P, and since we do not know exactly where our candidate's point is along $P'PP''$, we take refuge in guessing that he is the average man of his group and is at the point P itself. From P, therefore, we drop a perpendicular on to the vector x, and take the distance OX as representing his estimated score in that test. This geometrical procedure corresponds exactly to the calculation we made, as a little solid trigonometry will show the mathematical reader. The non-mathematical reader must take it on trust, but the model may illuminate the calculation. In our numerical example, taking the angles whose cosines are the correlations, the angle between y and z is about $72\frac{1}{2}°$, that between x and z is $60°$, and that between x and y about $46°$. It is worth the reader's while to draw y and z on a sheet of paper on the table, and to represent x by a knitting-needle rising at an angle above the table, making roughly angles of $46°$ with y and $60°$ with z. Any point P on the paper represents a person's scores in y and z, scores shared by all persons vertically above and below P. The projection of P on to the knitting-needle is $\hat{x}$, the estimate. It is the *average* of all the different scores x that a person with scores OY and OZ can have. The estimate will only be certain if the knitting-needle itself is on the table; it will be less and less certain, the more the knitting-needle is inclined to the table.

In Section 3 of Chapter IV we noted that the angles which three test vectors make with each other are impossible angles, if the determinant of the matrix of correlations becomes negative. Ordinarily, that determinant is positive. In our present example we have, for example:

$$\begin{vmatrix} 1{\cdot}0 & {\cdot}7 & {\cdot}5 \\ {\cdot}7 & 1{\cdot}0 & {\cdot}3 \\ {\cdot}5 & {\cdot}3 & 1{\cdot}0 \end{vmatrix} = {\cdot}38$$

Such a determinant, however, though it cannot be negative, can be zero, namely in the cases where the two smaller angles exactly equal the largest. In that case the

three vectors lie in one plane—the knitting-needle has sunk until it too lies on the table. In that case alone, when the determinant is zero, the "estimation" is certain, and all the people in the line $P'PP''$ have not only the same scores in y and z, but also the same scores in x. The vanishing of the above determinant therefore shows that this is so. And in more than three dimensions, although we can no longer make a model, the vanishing of the determinant:

$$\begin{vmatrix} 1 & r_{01} & r_{02} & r_{03} & . & r_{0n} \\ r_{01} & 1 & r_{12} & r_{13} & . & r_{1n} \\ r_{02} & r_{12} & 1 & r_{23} & . & r_{2n} \\ r_{03} & r_{13} & r_{23} & 1 & . & r_{3n} \\ . & . & . & . & . & . \\ r_{0n} & r_{1n} & r_{2n} & r_{3n} & . & 1 \end{vmatrix} = \Delta, \text{ say,}$$

shows that the criterion z_0 can be exactly estimated from the team $z_1, z_2 \ldots z_n$. In fact, the multiple correlation r_m, which we have already learned to calculate in another way, can also be calculated as—

$$r_m = \sqrt{1 - \frac{\Delta}{\Delta_{00}}}$$

where Δ is the whole determinant, and Δ_{00} is the minor left after deleting the criterion row and column. This expression clearly becomes equal to unity when $\Delta = 0$. In our small example x, y, z, we have—

$$\Delta = \cdot 38 \qquad \Delta_{00} = \cdot 91$$

$$r_m = \sqrt{1 - \frac{\cdot 38}{\cdot 91}} = \sqrt{\frac{\cdot 53}{\cdot 91}} = \sqrt{\cdot 5824} = \cdot 763$$

as we already know it to be from page 88.

9. *The "centroid" method and the pooling square.*—The pooling square, which we have learned to use in this chapter, enables us to see more clearly the nature of the factors first arrived at by Thurstone's "centroid" method. It will be remembered that in Chapter II, page 28, in a footnote we promised an explanation of this name "cen-

troid " (or centre of gravity) method as applied to the calculations of factor loadings.

Let us suppose that the tests z_1, z_2, z_3, and z_4 have the correlations shown, and let us by the aid of a pooling square find the correlation of each of them with the average of all. This means giving each test an equal weight in pooling it.

Equal Weights

		z_1	z_1	z_2	z_3	z_4
	z_1	1	1	r_{12}	r_{13}	r_{14}
Equal Weights	z_1	1	1	r_{12}	r_{13}	r_{14}
	z_2	r_{12}	r_{12}	1	r_{23}	r_{24}
	z_3	r_{13}	r_{13}	r_{23}	1	r_{34}
	z_4	r_{14}	r_{14}	r_{24}	r_{34}	1

The correlation of z_1 with the average of all is then obtained from the above pooling square, which condenses to:

1	$1 + r_{12} + r_{13} + r_{14}$
$1 + r_{12} + r_{13} + r_{14}$	*Sum of all the cells of the table of correlations.*

and the correlation coefficient is—

$$\frac{1 + r_{12} + r_{13} + r_{14}}{\sqrt{\text{above sum}}}$$

This, however, *is* exactly Thurstone's process applied to a table with full communalities of unity. The first Thurstone factor obtained from such a table is simply for each individual the average of his four test scores, and the method is called the " centroid " method, because " centroid " is the multi-dimensional name for an average (*Vectors*, Chapter III; and see Kelley, 1935, 59). The vector, in our geometrical picture, which represents the first Thurstone factor, is in the midst of the radiating

vectors which represent the tests, like the stick of a half-opened umbrella among the ribs. It does not, however, make equal angles with the test vectors unless these all make equal angles with each other. If several of them are clustered together, and the others spread more widely, the factor will lean nearer to the cluster. If in an extreme case we imagine several of the tests of the battery to be identical, and therefore with one identical vector, that vector would have to be weighted with the number of tests it represented; and a cluster of tests acts in somewhat the same way, and pulls the first-factor vector towards it. The position of balance which makes allowance for all the angular separations of the test vectors is not exactly the "central" position (unless they are quite symmetrically disposed), but the "centroid" position. And as we have seen, it corresponds to a straight average of each individual's *standardized* test scores, being itself then divided by its own standard deviation to standardize it.

In the foregoing explanation the communalities have been taken as unity, and the factor axis was pictured in the midst of the test vectors. If smaller communalities are used, the only difference is that a specific component of each test is discarded, and the first-factor axis must be pictured as in the midst of the vectors representing the other components of the tests. It can be shown that when communalities less than unity are used, if we bear in mind that the communal components of the tests are not then standardized, the pooling square gives the correlations with a weighted average exactly as before, except for the communalities instead of units in the diagonal. The average of the communal components therefore correlates with the first test thus:

1	h_1^2	r_{12}	r_{13}	r_{14}
h_1^2	h_1^2	r_{12}	r_{13}	r_{14}
r_{12}	r_{12}	h_2^2	r_{23}	r_{24}
r_{13}	r_{13}	r_{23}	h_3^2	r_{34}
r_{14}	r_{14}	r_{24}	r_{34}	h_4^2

1	$h_1^2 + r_{12} + r_{13} + r_{14}$
h_1^2 $+ r_{12}$ $+ r_{13}$ $+ r_{14}$	*Sum of all the cells in the table.*

which again gives Thurstone's loading for the first factor in the first test. His first factor is the average of the communal parts of the tests.

The later factors in their turn are, in a sense, averages of the residues. There are, however, some complications, the first being that the average of the residues just as they stand is zero. The manner in which Thurstone circumvents this has already been described in Chapter II.

CHAPTER VII

THE ESTIMATION OF FACTORS BY REGRESSION

1. *Estimating a man's "g."*—So far, our discussion of estimation in Chapter VI has had nothing immediate to do with factorial analysis. We are next, however, going to apply these principles of estimation to the problem of estimating a man's Spearman or Thurstone factors, given his test scores. As we have already explained in Chapter V, there is no need to "estimate" Hotelling's factors; they can be calculated without any loss of exactness because they are equal in number to the tests: and even if we analyse out only a few of them, they can be exactly calculated for a man from his test scores. When we say *exactly* here, we mean that the factors are known with the same exactness as the test scores which are our data.

Spearman or Thurstone factors, however, are more numerous than the tests, and can therefore only be "estimated." Two men with the same set of test scores may have different Thurstone factors. All we can do is to estimate them, and since the test scores of the two men are the same, our estimates of their most probable factors will be the same. The problem does not differ essentially from the estimation of occupational success or of ability in any "criterion" test. The loadings of a factor in each test give the z_0 row and column of the correlation matrix. Let us first consider the case of a hierarchical battery of tests, and the estimation of g, taking for our example the first four tests of the Spearman battery used as illustration in Chapter I, with these correlations:

	z_1	z_2	z_3	z_4
z_1	1·00	·72	·63	·54
z_2	·72	1·00	·56	·48
z_3	·63	·56	1·00	·42
z_4	·54	·48	·42	1·00

These correspond, in the analogy with the ordinary cases of estimation of the first part of this chapter, to the tests given to a candidate. In those cases, however, there was a real criterion whose correlations with the team of tests were known, and formed the z_0 row and column of the matrix. Here the "criterion" is g, and it cannot be measured directly; it can only be estimated in the manner we are now about to describe. We have here, therefore, no row and column of experimentally measured correlations for the criterion z_0 or g in the present case (Thomson, 1934*b*, 94). From the hierarchical matrix of inter-correlations of the tests, however, we can calculate the "saturation" or "loading" of each test with the hypothetical g, and use these for our criterion column and row of correlations. We thus arrive at the matrix:

	z_0	z_1	z_2	z_3	z_4
z_0	1·00	·90	·80	·70	·60
z_1	·90	1·00	·72	·63	·54
z_2	·80	·72	1·00	·56	·48
z_3	·70	·63	·56	1·00	·42
z_4	·60	·54	·48	·42	1·00

and we want to know the best-weighted combination of the test scores z_1 to z_4 in order to correlate most highly with $z_0 = g$. The problem is now the same as one of ordinary estimation of ability in an occupation, and the mathematical answer is the same. We can, for example, use Aitken's method of finding the regression coefficients, although in this case, because of the hierarchical qualities of the matrix, there is, as we shall shortly see, an easier method. It is, however, illuminating for the student actually to work out the regression coefficients as in an ordinary case of estimation, as shown on the next page.

If, therefore, we know the scores z_1, z_2, z_3, and z_4 which a man has made in these four tests, we can estimate his g by the equation (see overleaf)—

$$\hat{g} = \cdot 5531 z_1 + \cdot 2595 z_2 + \cdot 1602 z_3 + \cdot 1095 z_4$$

(1·00)	·72	·68	·54	−1·00	.	.	.	1·89
·72	1·00	·56	·48	.	−1·00	.	.	1·76
·68	·56	1·00	·42	.	.	−1·00	.	1·61
·54	·48	·42	1·00	.	.	.	−1·00	1·44
·90	·80	·70	·60	.	.	.	.	3·00
(2·0764)	(·4816)	·1064	·0912	·72	−1·00	.	.	·3992
	1·0000	·2209	·1894	1·495	−2·0764	.	.	·8289
	·1064	·6031	·0798	·68	.	−1·00	.	·4193
	·0912	·0798	·7084	·54	.	.	−1·00	·4194
	·1520	·1330	·1140	·90	.	.	.	1·2990
(1·7253)		(·5796)	·0596	·4709	·2209	−1·00	.	·3311
		1·0000	·1028	·8124	·3811	−1·7253	.	·5712
		·0597	·6911	·4037	·1894	.	−1·00	·3438
		·0994	·0852	·6728	·3156	.	.	1·1730
(1·4599)			(·6850)	·3552	·1666	·1030	−1·00	·3097
			1·0000	·5186	·2432	·1504	−1·4599	·4521
			·0750	·5920	·2777	·1715	.	1 1162
				·5531	·2595	·1602	·1095	1·0823

Regression Coefficients

The multiple correlation of such estimates in a large number of cases with the true values of g will be by analogy with our former case given by—

$$r_m^2 = \cdot 5531 \times \cdot 90 + \cdot 2595 \times \cdot 80 + \cdot 1602 \times \cdot 70 + \cdot 1095 \times \cdot 60 = \cdot 883$$

$$r_m = \cdot 940$$

We must remember, however, that such a correlation here is rather a fiction. We had in the former case the possibility of comparing our estimates with the candidate's eventual performance in the occupation or criterion z_0. Here we have no way of knowing g; we only have the estimates.

As before, we can check the whole calculation by a pooling square, thus:

		·5531	·2595	·1602	·1095
	1·00	·90	·80	·70	·60
·5531	·90	1·00	·72	·63	·54
·2595	·80	·72	1·00	·56	·48
·1602	·70	·63	·56	1·00	·42
·1095	·60	·54	·48	·42	1·00

Multiplying by the row weights and summing the columns condenses this to:

	·5531	·2595	·1602	·1095
1·000	·90	·80	·70	·60
·883	·900	·800	·700	·600

and multiplying by the column weights gives:

1·000	·883
·883	·883

showing that our calculation was exact to three places.

Estimating g from a hierarchical battery is therefore, mathematically, exactly the same problem as estimating any criterion, and can be done arithmetically in the same way. Because of the special nature of the hierarchical matrix of correlations, however, with its zero tetrad-differences, there is an easier way of calculating the estimate of g, due to Professor Spearman himself (*Abilities*, xviii). For its equivalence mathematically to the above see Thomson (1934*b*, 94–5) and Appendix, paragraph 10.

Meanwhile we shall illustrate it by an example which will at least show that it is equivalent in this instance. The calculation is best carried out in tabular form, and is based entirely on the saturations or loadings of the tests with g, which are also their correlations with g.

Test	r_{ig}	r_{ig}^2	$1 - r_{ig}^2$	$\frac{r_{ig}^2}{1 - r_{ig}^2}$	$\frac{r_{ig}}{1 - r_{ig}^2}$	Regression Coefficients $\frac{1}{1+S} \times \frac{r_{ig}}{1 - r_{ig}^2}$
1	·9	·81	·19	4·2632	4·7368	·5533
2	·8	·64	·36	1·7778	2·2222	·2596
3	·7	·49	·51	·9608	1 3725	·1603
4	·6	·36	·64	·5625	·9375	·1095

$$S = 7{\cdot}5643$$
$$1 + S = 8{\cdot}5643$$
$$\frac{1}{1+S} = {\cdot}1168$$

The result, with much less calculation, is the same. The quantity S is of some importance in this formula. It is formed in the fourth column of the table, from which it will be seen that—

$$S = \frac{r_{1g}^2}{1 - r_{1g}^2} + \frac{r_{2g}^2}{1 - r_{2g}^2} + \ldots = \Sigma \frac{r_{ig}^2}{1 - r_{ig}^2}$$

It is clear that S will become larger and larger as the number of tests is increased.

Now, we saw that the square of the multiple correlation r_m is obtained when we multiply each of the weights by r_{ig} and sum the products. That is to say—

$$\begin{aligned} r_m^2 &= \Sigma\,(\text{weight} \times \text{saturation}) \\ &= \Sigma \left(\frac{1}{1+S} \cdot \frac{r_{ig}}{1 - r_{ig}^2} \times r_{ig} \right) \\ &= \frac{1}{1+S} \cdot \Sigma \frac{r_{ig}^2}{1 - r_{ig}^2} = \frac{S}{1+S} \end{aligned}$$

This fraction will be the nearer to unity, the larger S is; and we can make S larger and larger by adding more and more (hierarchical) tests to the team. Thus in theory we can make a team to give as high a multiple correlation with g as we desire. It will also be noticed, however, from our table that the tests with high g saturation make

much the largest contribution to S, and therefore to the multiple correlation (see Piaggio, 1933, 89).

2. *Estimating two common factors simultaneously.*—We have seen in the preceding section how to estimate a man's g from his scores in a hierarchical team of tests, and in this we shall consider the broader question of estimating factors in general. Thus in Chapter II the four tests with correlations :

	1	2	3	4
1	.	·4	·4	·2
2	·4	.	·7	·3
3	·4	·7	.	·3
4	·2	·3	·3	.

were analysed into two common factors and four specifics with the loadings (see Chapter II, page 36).

	Common Factors		*Specific Factors*			
	I	*II*				
1	·5164	.	·8563	.	.	.
2	·7746	3162	.	·5477	.	.
3	7746	·3162	.	.	·5477	.
4	·3873	.	.	.	.	·9220

Any one column of these loadings can be used as the criterion row in the calculation by Aitken's method, and the regression coefficients calculated with which to weight a man's test scores in order to estimate that factor for him. If, as is probable, we want to estimate both common factors, we can do the two calculations together, as shown at top of next page. *Both* rows of loadings are written below the matrix of intercorrelations, and then pivotal condensation automatically gives both sets of regression coefficients, with only one extra row in each slab of the calculation, as on the next page.

If, therefore, we have a man's scores (in standard measure) in these four tests, our estimate of his Factor I will be (see overleaf)—

$$\cdot1787z_1 + \cdot3932z_2 + \cdot3932z_3 + \cdot1156z_4$$

(1·0)	·4	·4	·2	−1·0	.	.	.	1·0
·4	1·0	·7	·3	.	−1·0	.	.	1·4
·4	·7	1·0	·3	.	.	−1·0	.	1·4
·2	·3	·3	1·0	.	.	.	−1·0	·8
·5164	·7746	·7746	·3873	.	.	.	.	2·5429
.	·3162	·3162	.	.	.	.	.	·6324
	(·84)	·54	·22	·40	−1·0	.	.	1·0
	1·00	·6429	·2619	·4762	−1·1905	.	.	1·1905
	·54	·84	·22	·40	.	−1·0	.	1·0
	·22	·22	·96	·20	.	.	−1·0	·6
	·5680	·5680	·2840	·5164	.	.	.	1·9365
	·3162	·3162	.	.	.	.	.	·6324
		(·4928)	·0786	·1429	·6429	−1·0	.	·3571
		1·0000	·1595	·2900	1·3046	−2·0292	.	·7246
		·0786	·9024	·0952	·2619	.	−1·0000	·3381
		·2028	·1352	·2459	·6762	.	.	1·2608
		·1129	−·0828	−·1506	·3764	.	.	·2560
			(·8899)	·0724	·1594	·1594	−1·0000	·2811
			1·0000	·0814	·1791	·1791	−1·1237	·3159
			·1029	·1871	·4116	·4116	.	1·1134
			−·1008	−·1833	·2291	·2291	.	·1742
				·1787	·3932	·3932	·1156	1·0809
				−·1751	·2472	·2472	−·1133	·2060

Regression Coefficients

and estimates made in this way will have a multiple correlation r_m with the "true" values of the factor, in a number of different candidates, given by—

$$r_m^2 = \cdot 1787 \times \cdot 5164 + \cdot 3932 \times \cdot 7746 + \cdot 3932 \times \cdot 7746 + \cdot 1156 \times \cdot 3873 = \cdot 7462$$

$$\therefore r_m = \cdot 864$$

Similarly, the multiple correlation of the estimate of the second factor with the "true" values can be found to be—

$$r_m = \cdot 895$$

The two factors are not, therefore, estimated with equal accuracy by the team. As with ordinary estimation, the whole calculation can be checked by a pooling square. This check for the second factor is as follows:

		−·1751	·2472	·2472	−·1133
	1·0000	.	·3162	·3162	.
−·1751	.	1·0	·4	·4	·2
·2472	·3162	·4	1·0	·7	·3
·2472	·3162	·4	·7	1·0	·3
−·1133	.	·2	·3	·3	1·0

Multiplying the rows gives:

	−·1751	·2472	·2472	−·1133	
1·00000	.	·3162	·3162	.	
.	−·17510	−·07002	−·07004	−·03502	
·07816	·09888	·24720	·17304	·07416	
·07816	·09888	·17304	·24720	·07416	
.	−·02266	−·03399	−·03399	−·11330	
·15633	.	·31621	·31621	.	Sums of columns

Multiplying then by the column multipliers, and adding, we get:

1·00000	·15633
·15633	·15633

where the equality of the three quadrants shows that our regression weights were correct: and the multiple correlation is $\sqrt{\cdot 15633} = \cdot 395$.

We have now found the regression equations for estimating the two common factors by treating each in turn as a "criterion." It is also possible to estimate a man's specific factors in the same way. Indeed, we might, in the calculation opposite, have written the loadings of the four specific factors as four more rows below the common-factor loadings in the first slab, i.e.:

$$\begin{array}{cccc|}
\cdot 8563 & . & . & . \\
. & \cdot 5477 & . & . \\
. & . & \cdot 5477 & . \\
. & . & . & \cdot 9220
\end{array}$$

and calculated their regression coefficients all in the one calculation. But it is easier to obtain the estimate of a man's specific by subtraction (compare *Abilities*, 1932 edition, page xviii, line 10). For example, we know that the second test score is made up as follows—

$$z_2 = \cdot 7746 f_1 + \cdot 3162 f_2 + \cdot 5477 s_2$$

where f_1 and f_2 are the man's common factors and s_2 his specific. We have estimated his f_1 and f_2, and we know his z_2; so we can estimate his s_2 from this equation. The estimates of *all* a man's factors, to be consistent with the experimental data, must satisfy this equation and similar equations for the other tests. If the estimate of the specific is actually made by a regression equation, just like the other factors, it will be found to satisfy this requirement.* From the estimates of *all* a man's factors, therefore, including any specifics, we can reconstruct his scores in the tests exactly. From only a few factors, however, even from all the common factors, we cannot reproduce the scores exactly, but only approximately.

3. *An arithmetical short cut* (Ledermann, 1938*a*).—When the number of tests is *appreciably* greater than the number of common factors, the following scheme for computing the regression coefficients will involve less arithmetical labour than the general formulæ expounded in Chapter VI and applied to the factor problem in this chapter.

For illustration, we shall use the data of the preceding section (page 108), although in that example the number of tests (four) exceeds the number of common factors (two) only by two, which is too small an amount to demonstrate

* It is interesting to note that we know the best *relative* loadings of the tests to estimate a specific by regression without needing to know how many common factors there are, or whether indeed any specific exists or not. (Wilson, 1934. For the same fact in more familiar notation, see Thomson, 1936*a*, 43.)

fully the advantages of the present method. The common-factor loadings and the specifics of the four tests form a 4×2 matrix and a 4×4 matrix respectively, thus:

$$M_0 = \begin{bmatrix} \cdot 5164 & \cdot \\ \cdot 7746 & \cdot 3162 \\ \cdot 7746 & \cdot 3162 \\ \cdot 3873 & \cdot \end{bmatrix} ; \quad M_1 = \begin{bmatrix} \cdot 8563 & & & \\ & \cdot 5477 & & \\ & & \cdot 5477 & \\ & & & \cdot 9220 \end{bmatrix}$$

the matrix M_0 being identical with the first two columns, and the matrix M_1 with the last four columns of the table on page 107. Before the data are subjected to the computational routine process, which will again consist in the pivotal condensation of a certain array of numbers, some preliminary steps have to be taken: (i) the loadings of each test are divided by the square of its specific, and the modified values are then listed in a new 4×2 matrix:

$$M_0' = \begin{bmatrix} \cdot 7042 & \cdot \\ 2\cdot 5820 & 1\cdot 0540 \\ 2\cdot 5820 & 1\cdot 0540 \\ \cdot 4556 & \cdot \end{bmatrix}$$

e.g.
$$2\cdot 5820 = (\cdot 7746) \div (\cdot 5477)^2$$
$$1\cdot 0540 = (\cdot 3162) \div (\cdot 5477)^2$$

(ii) Next, the inner products (see footnote on page 31) of every column of M_0 in turn with every column of M_0' are calculated and arranged in a 2×2 matrix:

$$J = \begin{bmatrix} 4\cdot 5401 & 1\cdot 6329 \\ 1\cdot 6329 & \cdot 6665 \end{bmatrix}$$

i.e. the *first* row of this matrix contains the inner products of the *first* column of M_0 with all the columns of M_0', similarly the *second* row of J contains all those inner products which involve the *second* column of M_0, e.g.—

$$4\cdot 5401 = \cdot 5164 \times \cdot 7042 + \cdot 7746 \times 2\cdot 5820 + \cdot 7746 \times 2\cdot 5820 + \cdot 3873 \times \cdot 4556$$
$$1\cdot 6329 = \cdot 3162 \times 1\cdot 0540 + \cdot 3162 \times 1\cdot 0540$$

If there had been r common factors the matrix J would have been an $r \times r$ matrix. The arithmetic is simplified

by the fact that J is always symmetrical about its diagonal, so that only the entries on and above (below) the diagonal need be calculated. (iii) Finally, each element on the diagonal of J is augmented by unity, giving, in the notation of matrix calculus, the matrix:

$$I + J = \begin{vmatrix} 5\cdot5401 & 1\cdot6329 \\ 1\cdot6329 & 1\cdot6665 \end{vmatrix}$$

This matrix is now "bordered" below by the matrix M_0', and on the right-hand side by a block of minus ones and zeros in the usual way. The process of pivotal condensation then yields the same regression coefficients as were obtained on page 108.

5·5401	1·6329	− 1·0000	·	6·1730
1·0000	·2947	− ·1805	·	1·1142
1·6329	1·6665	·	− 1·0000	2·2994
·7042	·			·7042
2·5820	1·0540			3·6360
2·5820	1·0540			3·6360
·4556	·			·4556
	1·1853	·2947	− 1·0000	·4800
	1·0000	·2486	− ·8437	·4050
	− ·2075	·1271		− ·0804
	·2931	·4661		·7591
	·2931	·4661		·7591
	− ·1343	·0822		− ·0520
		·1787	− ·1751	·0036
Regression Coefficients		·3932	·2473	·6404
		·3932	·2473	·6404
		·1156	− ·1133	·0023

4. *Reproducing the original scores.*—Let us imagine a man who in each of the four tests in our example obtains a score of $+ 1$; that is, one standard deviation above the average. We choose this set of scores merely to make the arithmetic of the example easy. The regression estimates of his two common factors are—

$$\hat{f}_1 = \cdot1787z_1 + \cdot3932z_2 + \cdot3932z_3 + \cdot1156z_4$$
$$\hat{f}_2 = - \cdot1751z_1 + \cdot2472z_2 + \cdot2472z_3 - \cdot1133z_4$$

Inserting his scores $z_1 = z_2 = z_3 = z_4 = 1$ into these equations we get for the regression estimates of his factors—

$$\hat{f}_1 = 1{\cdot}0807$$
$$\hat{f}_2 = \cdot 2060$$

that is, we estimate his first factor to be rather more than one standard deviation, his second factor to be about one-fifth of a standard deviation, above the average.

Now, the specification equations which give the composition of the four tests in terms of the factors are—

$$\begin{aligned} z_1 &= \cdot 5164 f_1 \qquad\quad . \quad + \cdot 8563 s_1 \\ z_2 &= \cdot 7746 f_1 + \cdot 3162 f_2 + \cdot 5477 s_2 \\ z_3 &= \cdot 7746 f_1 + \cdot 3162 f_2 + \cdot 5477 s_3 \\ z_4 &= \cdot 3873 f_1 \qquad\quad . \quad + \cdot 9220 s_4 \end{aligned}$$

If we insert the above estimates $\hat{f}_1$ and $\hat{f}_2$ in lieu of f_1 and f_2, we get for this man's scores—

$$\begin{aligned} z_1 &= \cdot 5581 + \cdot 8563 \hat{s}_1 \\ z_2 &= \cdot 9022 + \cdot 5477 \hat{s}_2 \\ z_3 &= \cdot 9022 + \cdot 5477 \hat{s}_3 \\ z_4 &= \cdot 4186 + \cdot 9220 \hat{s}_4 \end{aligned}$$

We know his four scores each to have been $+ 1$, and if we had also worked out the estimates of his specifics by the regression method we should have found that they added just enough to the above equations to make each indeed come to $+ 1$. We can, therefore, find his estimated specifics more easily from the above equations, as in this case—

$$\hat{s}_1 = \frac{1 - \cdot 5581}{\cdot 8563} = \cdot 5161$$

$$\hat{s}_2 = \frac{1 - \cdot 9022}{\cdot 5477} = \cdot 1786$$

and so for $\hat{s}_3$ and $\hat{s}_4$, subtracting the contribution of the common factors from the known score (here $+ 1$ in each case) and dividing by the specific loading.

The *regression* estimates of the factors, made by the system we have so far been considering, are as a matter of fact not the only estimates which have been proposed. The alternative system has certain advantages, to be explained later. The regression estimates are the best in

the sense, as we said when deducing them, that they give the highest correlation, taken over a large number of men, between the estimates and the true values of a criterion when the latter can be separately ascertained. Just what this correlation means, however, when there is no possibility of ascertaining the "true" values (for factors, when they outnumber the tests, only can be estimated) it is not so easy to say.

The regression estimates of the factors, as calculated in the present chapter, have one other great advantage, that they are consistent with the ordinary estimation of vocational ability made without using factors at all, as can best be shown by means of the example of Section 7 of Chapter VI.

5. *Vocational advice with and without factors.*—In that example we had an "occupation" z_0, and four tests z_1, z_2, z_3, and z_4; and in Chapter VI, without using factors at all, we arrived at the following estimation of a man's success or "score" in the occupation (which is, after all, only a test like the others, though a long-drawn-out one)—

$$\dot{z}_0 = \cdot 390 z_1 + \cdot 222 z_2 + \cdot 018 z_3 + \cdot 431 z_4$$

Now let us suppose that the matrix of correlations of these five tests (including the occupation as a test) had been analysed, by Thurstone's method or any other, into common factors and specifics—the matrix is given in Chapter VI, page 91. Indeed, the four tests proper were so analysed by Dr. Alexander in the monograph from which we took their correlations, and the analysis below is based on his. The "occupation" z_0 is a pure fiction made for the purpose of this illustration, but we can easily imagine it also being analysed in exactly the same way as a test. The table of loadings of the factors, to which we may as well give Dr. Alexander's names of g (Spearman's g), v (a verbal factor), and F (a practical factor), is as follows:

		g	v	F	*Specific*
Occupation	z_0	·55	·45	·60	·37
Stanford-Binet	z_1	·66	·52	·21	·50
Reading test	z_2	·52	·66	.	·54
Geometrical analogies	z_3	·74	.	.	·67
Picture completion	z_4	·37	.	·71	·60

With this table of loadings in our possession we might have given vocational advice to a man in a roundabout way. Instead of inserting his scores in z_1, z_2, z_3, and z_4 in the equation (see page 93).

$$\hat{z}_0 = \cdot 390z_1 + \cdot 222z_2 + \cdot 018z_3 + \cdot 481z_4$$

we might have estimated his factors g, v, and F from his scores in the four tests, and then inserted these estimated factors in the specification equation of the occupation—

$$z_0 = \cdot 55g + \cdot 45v + \cdot 60F + \cdot 37s_0$$

(ignoring the specific s_0, which cannot be estimated from z_1, z_2, z_3, and z_4). *Had we done so, we should have arrived at exactly the same numerical estimate of his z_0 as by the direct method* (Thomson, 1936*a*, 49 and 50).

The actual estimation of the factors g, v, and F from the four tests will form a good arithmetical exercise for the student. The beginning and end of the calculation of the regression coefficients is shown here, following exactly the lines of the smaller example on page 108 of this chapter:

								Check
1·00	·69	·49	·39	− 1	.	.	.	1·57
·69	1·00	·38	·19	.	− 1	.	.	1·26
·49	·38	1·00	·27		.	− 1	.	1·14
·39	·19	·27	1·00	.	.	.	− 1	·85
·66	·52	·74	·37	.		.	.	2·29
·52	·66	.	.	.	.	.	.	1·18
·21	.	.	·71	.	.	.	.	·92

This reduces by pivotal condensation step by step to the three sets of regression coefficients:

for $\hat{g}$	·300	·095	·532	·095
for $\hat{v}$	·353	·581	− ·352	− ·153
for $\hat{F}$	·121	− ·148	− ·206	·747

The result is to give us three equations for estimating g, v, and F from a man's scores in the four tests, viz.—

$$\hat{g} = \cdot 300z_1 + \cdot 095z_2 + \cdot 532z_3 + \cdot 095z_4$$
$$\hat{v} = \cdot 353z_1 + \cdot 581z_2 - \cdot 352z_3 - \cdot 153z_4$$
$$\hat{F} = \cdot 121z_1 - \cdot 148z_2 - \cdot 206z_3 + \cdot 747z_4$$

Now let us assume a set of scores z_1, z_2, z_3, z_4 for a man, and see what the estimate of his occupational ability is by

the two methods, the one direct without using factors, the other by way of factors. Suppose his four scores are—

$$\begin{array}{cccc} z_1 & z_2 & z_3 & z_4 \\ \cdot 2 & -\cdot 4 & \cdot 7 & \cdot 6 \end{array}$$

The estimates of his factors g, v, and F will therefore be—

$$\begin{aligned} \hat{g} &= \cdot 300 \times \cdot 2 + \cdot 095 \times (-\cdot 4) + \cdot 532 \times \cdot 7 + \cdot 095 \times \cdot 6 = \quad \cdot 451 \\ \hat{v} &= \cdot 353 \times \cdot 2 + \cdot 581 \times (-\cdot 4) - \cdot 352 \times \cdot 7 - \cdot 153 \times \cdot 6 = -\cdot 500 \\ \hat{F} &= \cdot 121 \times \cdot 2 - \cdot 148 \times (-\cdot 4) - \cdot 206 \times \cdot 7 + \cdot 747 \times \cdot 6 = \quad \cdot 387 \end{aligned}$$

If now we insert these estimates of his factors into the specification equation of the occupation, ignoring its specific, we get for our estimate of his occupational success :

$$\hat{z}_0 = \cdot 55 \times \cdot 451 + \cdot 45 \times (-\cdot 500) + \cdot 60 \times \cdot 387 = \cdot 255$$

that is, we estimate that he will be about a quarter of a standard deviation better than the average workman. This by the indirect method using factors.

By the direct method, without using factors at all, we simply insert his test scores into the equation—

$$\hat{z}_0 = \cdot 390 z_1 + \cdot 222 z_2 + \cdot 018 z_3 + \cdot 431 z_4$$

and obtain—

$$\begin{aligned} \hat{z}_0 &= \cdot 390 \times \cdot 2 + \cdot 222 \times (-\cdot 4) + \cdot 018 \times \cdot 7 + \cdot 431 \times \cdot 6 \\ &= \cdot 260 \end{aligned}$$

exactly the same estimate as before—for the difference in the third decimal place is entirely due to " rounding off " during the calculation. The third decimal place of the direct calculation is more likely to be correct, since it is so much shorter.

6. *Why, then, use factors at all ?*—The reader may now ask, " What, then, is the use of estimating a man's factors at all ? " Well, in a case analogous to that of the present example, it is quite unnecessary to use factors at all, and there is no doubt that a great many experimenters have rushed to factorial analysis with quite unjustifiable hopes of somehow getting more out of it than ordinary methods of vocational and educational advice can give without mentioning factors. But we must not go to the other extreme and " throw out the baby with the bath-water." There may be other reasons for using factors, apart from

vocational advice. And even in giving such advice, which really means describing men and occupations in similar terms, so that we can see if they fit one another or not, it may be that factors have some advantages not disclosed by the above calculation.

This man whom we have used above, for example, may be described either in terms of his scores in four fairly well-known tests, or in terms of the factors g, v, and F. By the former method his description is:

Stanford-Binet test	·2, slightly above average
Thorndike reading test	—·4, distinctly below average
Spearman's geometrical analysis . . .	·7, good
Picture-completion test	·6, good

This description already suggests to us that he is a man of average intelligence or rather better, of not much schooling, and with a bit of a gift for seeing shapes, and similarities in them. From the correlations of the occupation with these four tests we know that it most resembles the first and last tests and least resembles the third. We can probably draw the conclusion that this man will be above average in it; and we can draw this conclusion accurately if we calculate the regression equation—

$$\dot{z}_0 = \cdot 390z_1 + \cdot 222z_2 + \cdot 018z_3 + \cdot 431z_4$$

As a description of the man, however, the above table suffers from the fact that the four tests are correlated with one another. We feel a certain clarity in the description in terms of factors, because these are independent of one another and uncorrelated. This man whom we are at present considering is alternatively described, in terms of factors, as:

Factor	*Estimated Amount*
g	·451
v	—·500
F	·387

that is, a quite intelligent (g) and practical (F) man with, however, not much ability in using and understanding words (v). There is a certain air of greater generality about the factors than there is about the particular tests

from which they have been deduced, and they give definition and point to mental descriptions, or at least they seem to do so.

Yet some of these " advantages " of using factors begin to look less bright when looked into more carefully. We said that one advantage is that factors are independent and uncorrelated. So they are, if their true values are known. *But we only know their estimates, and these are correlated,* as we shall illustrate shortly. If we use factors it is clear that we must, if we value the advantage of independence, seek to obtain estimates which are as little correlated with one another as possible. There have been proposals to use factors which are *really* correlated; not merely correlated when their estimates are taken, but correlated in their true measures. What advantage can these have over the actual correlated tests? The fundamental advantage hoped for by the factorist seems to be that the factors (correlated or uncorrelated) may turn out to be comparatively few in number, and may thus replace a multitude of tests and innumerable occupations by a description in these few factors. The student whose knowledge of the subject is being obtained from this book is not yet equipped to discuss adequately the very fundamental questions raised in this section, to which we shall return several times in later chapters. One last point in favour of factors may, however, be expanded somewhat here. We said a couple of sentences back that factorists hope to give adequate descriptions of men and of occupations in terms of a comparatively small number of factors. This, if achieved, would react on social problems somewhat in the same way as the introduction of a coinage influences trade previously carried on by barter. A man can exchange directly five cows for so many sheep, so much cloth, and a new ploughshare; but the transaction is facilitated if each of these articles is priced in pounds, shillings, and pence, or in dollars and cents, even though the end result is the same. And so perhaps with the " pricing " of each man and each occupation in terms of a few factors.

But the prices must be accurate; and the analyses of

tests and occupations into factors, still more the calculation of quantitative estimates of these factors, are as yet very inaccurate, and perhaps are inherently subject to uncertainty. A fluctuating and doubtful coinage can be a positive hindrance to trade, and barter may be preferable in such circumstances.

We showed in Section 5 above that a direct regression estimate of a man's ability in an occupation gives identically the same result as an estimate via the roundabout path of factors, so that at least when the direct regression estimate is possible there can be no quantitative advantage in using factors. When, however, is the direct regression estimate possible, and when is it impossible?

To make the direct regression estimate we require the complete table of correlations of the tests with one another *and with the occupation*, and we have to know the candidate's scores in the tests. This implies that these same tests have been given to a number of workers whose proficiency in the occupation is known, for otherwise we would not know the correlations of the tests with the occupation. Under these ideal circumstances any talk of factors is certainly unnecessary so far as obtaining a quantitative estimate is concerned.

But suppose these ideal conditions do not hold! These tests which we have given to the candidate have never been given, at any rate as a battery, to workers in the occupation, and their correlations with the occupation are unknown! This situation is particularly likely to arise in vocational advice or guidance as distinguished from vocational selection. In the latter we are, usually on behalf of the employer, selecting men for a particular job, and we are practically certain to have tried our tests on people already in the job, and to be in a position to make a direct estimation without factors. But in vocational guidance we wish to gauge the young person's ability in very many occupations, and it is unlikely that just this battery of tests that we are using has been given to workers in all these different jobs. In that case we cannot make a direct regression estimate of our candidate's probable proficiency in every occupation. Can we, then, obtain an estimate in any other way?

Other ways are conceivable, but it must at the outset be emphasized that *they are bound to be less accurate than the direct estimate without factors.* Although this battery of tests has not been given to workers in the occupation, perhaps other tests have, and by the aid of that other battery a factor analysis of the occupation has perhaps been made. If our tests enable the same factors to be estimated, we can gauge the man's factors and thence indirectly his occupational proficiency. Unfortunately, the "if" is a rather big one. Are factors obtained by the analysis of different batteries of tests the same factors; may they not be different even though given the same name? We shall discuss this very important point later, but meanwhile let us suppose that we have reasonable confidence in the identity of factors called by the same name by different workers with different batteries. Then the probable course of events would be something like this. An experimenter, using whatever tests he thinks practicable and suitable, analyses an occupation into factors. Another experimenter, at a different time and place, is asked to give advice to a candidate for that occupation. Using whatever tests he in his turn has available, he assesses in this candidate the factors which the previous experimenter's work leads him to think are necessary in the occupation, and gives his advice accordingly. The factors have played their part as a go-between, like a coinage. All depends on the confidence we have in the identity of the factors. We shall see later that there is only too much reason to think that the possibility of this confidence being misplaced has hardly been sufficiently realized by many over-enthusiastic factorists. And even if the common factors are identical, there remains the danger that the "specific" of the occupation may be correlated with some of the "specifics" of the tests, a fact which cannot be known unless the same tests have been given to workers in the occupation.

7. *The geometrical picture of correlated estimates.*—Of the swarm of difficulties and doubts raised by these remarks we shall choose one to deal with first. We said that even although we make our analysis of the tests we use into *uncorrelated* factors, the estimates of these factors will be

correlated. This can best be appreciated if we consider what the estimation of factors means in terms of the geometrical picture of Chapter IV, which we also used in Chapter VI, Figure 19 (page 96). In this latter figure we were illustrating the straightforward process of estimating a " criterion " x, given a man's scores in two tests y and z. We saw that these two scores did not tell us exactly the man's position in the three-dimensional space of x, y, and z, but only told us that he stood somewhere along a line $P'PP''$ at right angles to the plane of yz. In default of his exact point, we took the point P, which is where the average man of the array $P'PP''$ stands, and by projection from it on to the vector x found an estimate OX of his x score.

Exactly the same picture will serve for the estimation of a factor, if we suppose the vector x to be now the vector of a factor (say a) whose angles with y and z are known—for the loadings of y and z with a are their cosines.

Now, suppose that we are referring these two tests y and z to three uncorrelated factors. It is immaterial whether any of these factors are specifics, for a specific is estimated exactly like any other factor. We shall call them simply a, b, and c. Since the three factors are uncorrelated, they are represented in the geometrical picture by orthogonal (i.e. rectangular) axes, as shown in Figure 20. The vectors a and b are at right angles to each other in the plane of the paper, while the vector c is at right angles to both of them, standing out from the paper. These axes are to be imagined as continued backwards in their negative directions also, but only their positive portions are shown, to avoid confusing the diagram.

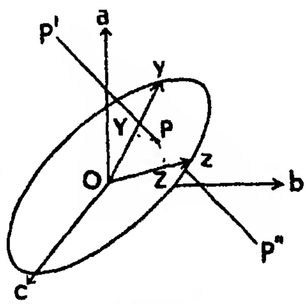

Figure 20.

The vectors y and z, also shown only in their positive directions, represent the two tests, and the angle between them represents by its cosine their correlation with one another. These two vectors y and z are not in

any of the planes *ab*, *bc*, or *ca*, but project into the space between them.

The three orthogonal planes *ab*, *bc*, and *ca* divide the whole of three-dimensional space into eight octants, and if as is usual the final positions chosen for *a*, *b*, and *c* are such that all loadings are positive, the positive directions of *y* and *z* will project into the positive octant as shown in the figure, in which the vector *z* is coming out of the paper more steeply than *y* is.

The two vectors *Oy* and *Oz* define a plane, on which a circle has been drawn, which in the figure appears as an ellipse, since the plane *yz* is not in the paper but inclined to it.

In the three-dimensional space defined by *abc*, the population of all persons is represented by a spherical swarm of points dense at *O*, more sparse as the distance from *O* increases in any direction. From any point in this space, perpendiculars can be dropped to *a*, *b*, *c*, *y*, and *z*, and the distances from *O* to the feet of these perpendiculars represent the amount of the factors *a*, *b*, and *c* possessed by the person whom that point represents, and his scores in the two tests. Conversely, a knowledge of his three factors would enable us to identify his point by erecting three perpendiculars and finding their meeting-point. But a knowledge of his scores in *y* and *z* does *not* enable us to identify his point, but only to identify a line *P′PP″*, anywhere on which his point may lie. In the figure, let *OY* and *OZ* represent a person's scores in *y* and *z*. Then on the plane *yz* we may draw perpendiculars meeting at *P*. But the point representing the person whose scores are *OY* and *OZ* need not be at *P*; it can be anywhere in *P′PP″* at right angles to the plane *yz*, for wherever it is on this line, perpendiculars from it on to *y* and *z* will fall on the points *Y* and *Z*. In estimating factors from tests we have to choose one point on *P′PP″* from which to drop perpendiculars on to *a*, *b*, and *c*, and we choose *P* because the man at *P* is the average man of the array of men *P′PP″*. Thus when we are estimating factors, all our population is represented by points on the plane *yz* (the plane on which in the figure the circle is drawn which

looks like an ellipse), although really they should be represented by a spherical swarm of dots.

When the population is truly represented by its spherical swarm of dots, the axes *a*, *b*, and *c* represent uncorrelated factors. But when the spherical swarm of dots has been collapsed or projected on to the diametrical plane *yz* this is no longer the case. By taking only points in the plane *yz* from which to estimate factors in a three-dimensional space we have passed as it were from the geometrical picture of Chapter IV to the geometrical picture used in the first portion of Chapter V, where correlation between rectangular axes was indicated by an ellipsoidal distribution of the population points. We have introduced correlation between the estimates of *a*, *b*, and *c*, because we have distorted the distribution of the population from a three-dimensional sphere to a flat circle on the plane *yz*, that is to an ellipsoid, for in a space of three dimensions the circle is an ellipsoid with two axes equal and the third one zero. Consider, for example, the particular point *P* shown in the figure. From it, projections on to *a*, *b*, and *c* are all positive, the man with scores *OY* and *OZ* in *y* and *z* is estimated to have all three factors above the average, which adds to their positive correlation. But in actual fact, since *P* may really lie anywhere along *P'PP"*, a line which does not remain for its whole length in the positive octant *abc*, the man may really have some of his factors positive and some negative.

If, together with the population, the rectangular axes *a*, *b*, and *c* are also projected on to the plane *yz*, these projections will not all be at right angles—obviously they cannot, for three lines *in a plane* cannot all be at right angles to one another. *The angles between these projections* of the factor axes on to the test plane *represent the correlations between the estimated factors.*

Our illustration has been only in two and three dimensions, for clearness and to permit of figures being drawn. Similar statements, however, are true of more tests and more factors, where the spaces involved are of dimensions higher than three. If there are *n* tests, the *n* test vectors define an *n*-space, analogous to the *yz* plane of Figure 20.

If these n tests have been analysed into r common factors and n specifics, $n + r$ factors in all, the factor axes will define an $(n + r)$ space analogous to the three-dimensional *abc* space of Figure 20. A man's n scores in the tests define his position P in the n-space of the tests, but he may be anywhere in a *space* $P'PP''$, of r dimensions, at right angles to the test space, analogous to the *line* $P'PP''$ in Figure 20. We take the point P to represent him *faute de mieux*, and project the distance OP on to the factor axes to get his estimated factors. These estimated factors are correlated with one another, and if we project the $n + r$ factor axes from the $(n + r)$-space on to the n-space of the tests, the angles between these shadow vectors represent the correlations between the estimates.

8. *Calculation of correlation between estimates.*—Arithmetically, these correlations are easily calculated from the inner products of (*b*), the loadings of the estimated factors with the tests (page 115), with (*a*), the loadings of the tests with the factors (page 114). Moreover, this gives us the opportunity to explain in passing what is meant by "matrix multiplication."

The matrix of loadings of the four tests with the three common factors is (page 114):

$$M = \begin{vmatrix} \cdot66 & \cdot52 & \cdot21 \\ \cdot52 & \cdot66 & . \\ \cdot74 & . & . \\ \cdot37 & . & \cdot71 \end{vmatrix}$$

and the matrix of the loadings of the three estimated factors with the four tests is (page 115):

$$N = \begin{vmatrix} \cdot300 & \cdot095 & \cdot532 & \cdot095 \\ \cdot358 & \cdot581 & -\cdot352 & -\cdot158 \\ \cdot121 & -\cdot148 & -\cdot206 & \cdot747 \end{vmatrix}$$

Then the matrix of variances and covariances of the estimated factors is—

$$K = NM$$

in which formula we must explain how we form the product of two matrices. By the product of two matrices

we mean the new matrix formed of the inner products of the rows of the left-hand matrix with the columns of the right-hand matrix, set down in the order as formed. Thus, in forming the product:

$$NM = \begin{bmatrix} \cdot 300 & \cdot 095 & \cdot 532 & \cdot 095 \\ \cdot 353 & \cdot 581 & -\cdot 352 & -\cdot 153 \\ \cdot 121 & -\cdot 148 & -\cdot 206 & \cdot 747 \end{bmatrix} \begin{bmatrix} \cdot 66 & \cdot 52 & \cdot 21 \\ \cdot 52 & \cdot 66 & . \\ \cdot 74 & . & . \\ \cdot 37 & . & \cdot 71 \end{bmatrix}$$

$$= \begin{matrix} \cdot 676 & \cdot 219 & \cdot 130 \\ \cdot 218 & \cdot 567 & -\cdot 034 \\ \cdot 127 & -\cdot 035 & \cdot 556 \end{matrix} = K$$

the first element ·676 of K is the inner product of the first *row* of N with the first *column* of M—

$$\cdot 300 \times \cdot 66 + \cdot 095 \times \cdot 52 + \cdot 532 \times \cdot 74 + \cdot 095 \times \cdot 37 = \cdot 676$$

In the same way, every element in K is formed. The element $-$ ·034, in the second row and third column of K, is the inner product of the second row of N with the third column of M—

$$\cdot 353 \times \cdot 21 + \cdot 581 \times \text{zero} - \cdot 352 \times \text{zero} - \cdot 153 \times \cdot 71 = - \cdot 034$$

If our arithmetic throughout the whole calculation of these loadings had been perfectly accurate, the matrix K would have been perfectly symmetrical about its diagonal. The actual discrepancies (as ·127 and ·130) are a measure of the degree of arithmetical accuracy attained.*

The matrix K thus arrived at gives by its diagonal elements ·676, ·567, and ·556, the *variances* of the three estimated factors (that is, the squares of their standard deviations), and by its other elements their *covariances* in pairs (that is, their overlap with one another). The correlation of any two estimated factors is equal to (see Chapter I, Figure 2)—

* A trial will show the reader that the product NM is quite different from the product MN. This is the only fundamental difference between matrix algebra and ordinary algebra.

$$r_{ij} = \frac{\text{covariance } (ij)}{\sqrt{\text{variance } (i) \times \text{variance } (j)}}$$

From K we can therefore form the matrix of correlations of the estimated factors. It is:

$$\begin{bmatrix} 1{\cdot}000 & {\cdot}353 & {\cdot}212 \\ {\cdot}353 & 1{\cdot}000 & -{\cdot}061 \\ {\cdot}212 & -{\cdot}061 & 1{\cdot}000 \end{bmatrix}$$

wherein ·353, for example, is $\cdot 219 \div \sqrt{(\cdot 676 \times \cdot 567)}$. Although, therefore, the "true" factors g and v are uncorrelated, their estimates g and v are correlated to an amount ·353. The "true" factors g, v, and F are in standard measure, but their estimates g, v, and F have variances of only ·676, ·567, and ·556 instead of unity. These variances, be it noted in passing, are equal also to the squares of the correlations between g and $\hat{g}$, v and $\hat{v}$, F and $\hat{F}$.

Not only are the estimates of the common factors correlated among themselves; they are correlated with the specifics, so that the *estimates* of the specifics are not strictly specific. As a numerical illustration we may take the hierarchical matrix used in Section 1, pages 102 *ff.*, four tests of the larger hierarchical matrix used in Chapters I (page 6) and II (page 23).

	z_1	z_2	z_3	z_4
z_1	1·00	·72	·63	·54
z_2	·72	1·00	·56	·48
z_3	·63	·56	1·00	·42
z_4	·54	·48	·42	1·00

The regression estimate of g from this battery is, as we found on page 104)—

$$\hat{g} = \cdot 553 z_1 + \cdot 259 z_2 + \cdot 160 z_3 + \cdot 109 z_4$$

The regression estimates for the four specifics can also be found, either by a full calculation like that of page 108, or by the simpler method of subtraction of page 110. Thus, to estimate s_1 in our present example we know that—

$$z_1 = \cdot 9g + \sqrt{1 - \cdot 9^2}\, s_1$$
$$= \cdot 9g + \cdot 436 s_1$$

Also we know that the estimates $\hat{g}$ and $\hat{s}_1$ will satisfy the same equation—

$$z_1 = \cdot 9\hat{g} + \cdot 436\hat{s}_1$$

that is—

$$\hat{s}_1 = \frac{z_1 - \cdot 9\hat{g}}{\cdot 436}$$

On inserting the expression for $\hat{g}$ into this we get—

$$\hat{s}_1 = 1 \cdot 152z_1 - \cdot 535z_2 - \cdot 333z_3 - \cdot 225z_4$$

and similarly—

$$\hat{s}_2 = - \cdot 737z_1 + 1 \cdot 313z_2 - \cdot 215z_3 - \cdot 145z_4$$
$$\hat{s}_3 = - \cdot 542z_1 - \cdot 253z_2 + 1 \cdot 242z_3 - \cdot 106z_4$$
$$\hat{s}_4 = - \cdot 415z_1 - \cdot 194z_2 - \cdot 121z_3 + 1 \cdot 169z_4$$

We have now both N, the matrix of loadings of the *estimated* factors $\hat{g}$, $\hat{s}_1$, $\hat{s}_2$, $\hat{s}_3$, $\hat{s}_4$ with the four tests, and M, which we already know, the matrix of loadings of the four tests with the five factors g, s_1, s_2, s_3, and s_4, namely :

$$M = \begin{bmatrix} \cdot 9 & \cdot 436 & . & . & . \\ \cdot 8 & . & \cdot 600 & . & . \\ \cdot 7 & . & . & \cdot 714 & . \\ \cdot 6 & . & . & . & \cdot 800 \end{bmatrix}$$

From their product NM we obtain the matrix K of variances and covariances of the estimated factors, namely :

$$\begin{bmatrix} \cdot 553 & \cdot 259 & \cdot 161 & \cdot 109 \\ 1 \cdot 152 & - \cdot 535 & - \cdot 333 & - \cdot 225 \\ - \cdot 737 & 1 \cdot 313 & - \cdot 215 & - \cdot 145 \\ - \cdot 542 & - \cdot 253 & 1 \cdot 242 & - \cdot 106 \\ - \cdot 415 & - \cdot 194 & - \cdot 121 & 1 \cdot 169 \end{bmatrix} \begin{bmatrix} \cdot 9 & \cdot 436 & . & . & . \\ \cdot 8 & . & \cdot 600 & . & . \\ \cdot 7 & . & . & \cdot 714 & . \\ \cdot 6 & . & . & . & \cdot 800 \end{bmatrix}$$

$$= \begin{bmatrix} \cdot 880 & \cdot 241 & \cdot 155 & \cdot 115 & \cdot 087 \\ \cdot 241 & \cdot 502 & - \cdot 321 & - \cdot 238 & - \cdot 180 \\ \cdot 150 & - \cdot 321 & \cdot 788 & - \cdot 154 & - \cdot 116 \\ \cdot 116 & - \cdot 236 & - \cdot 152 & \cdot 887 & - \cdot 085 \\ \cdot 088 & - \cdot 181 & - \cdot 116 & - \cdot 086 & \cdot 935 \end{bmatrix} = K$$

Again, we have a check on the accuracy of our arithmetic, for K will, if we have been accurate, be exactly symmetrical about its principal diagonal, i.e. its diagonal running from north-west to south-east. The largest discrepancy in our case is between ·150 and ·155. Moreover, since in this case K includes *all* the factors, we have another check which was not available when we calculated a K for common factors only: the sum of the elements in the principal diagonal (called the "trace," or in German the "*Spur*") here must come out equal to the number of tests. In our case we have—

$$\cdot 880 + \cdot 502 + \cdot 788 + \cdot 887 + \cdot 935 = 3\cdot 992$$

and there are four tests. These elements which form the trace of K are, it will be remembered, the variances of the estimates $\hat{g}$, $\hat{s}_1$, $\hat{s}_2$, $\hat{s}_3$, and $\hat{s}_4$. So that we see that the total variances of the *five* factors is no greater than the total variance (viz. 4) of the *four* tests in standard measure. This is only another instance of the general law that we cannot get more out of anything than we put into it (at any rate, not in the long run).

From K we can at once calculate the correlation of the estimated factors. Adjusting the slight arithmetical departures from symmetry, we get:

	$\hat{g}$	$\hat{s}_1$	$\hat{s}_2$	$\hat{s}_3$	$\hat{s}_4$
$\hat{g}$	1·000	·362	·184	·131	·096
$\hat{s}_1$	·362	1·000	− ·510	− ·354	− ·263
$\hat{s}_2$	·184	− ·510	1·000	− ·183	− ·135
$\hat{s}_3$	·131	− ·354	− ·183	1·000	− ·094
$\hat{s}_4$	·096	− ·263	− ·135	− ·094	1·000

from which we see that $\hat{g}$ is correlated with each of the estimated specifics positively, while the latter are correlated negatively among themselves, in this (a hierarchical) example.

We have then this result, that although we set out to analyse our battery of tests into independent uncorrelated factors, the estimates which we make of these factors are correlated with one another, and instead of being in

standard measure have variances, and therefore standard deviations, less than unity. We could, of course, make them unity by dividing all our estimates by their calculated standard deviation. But that would make no change in their correlations.

The cause of all this is the excess of factors over tests, and consequently this drawback—the correlation of the estimates—depends upon the ratio of the number of factors to the number of tests. The extra factors are the common factors, for there is a specific to each test, and therefore with the same number of common factors the correlation between the estimates will decrease as the number of tests in the battery increases. Just as in the hierarchical case one of the tasks of the experimenter is to find tests to add to the number in his battery without destroying its hierarchical nature, so in the case of a Thurstone battery which can be reduced to rank 2, 3, 4 . . . or r, a task will be to add tests to the battery which with suitable communalities will leave the rank unchanged and the pre-existing communalities unaltered, in order that the common factors may be the more accurately estimated, and the estimates be more nearly uncorrelated.

With Thurstone batteries of tests, therefore, we arrive at the same necessity to "purify" any extended battery as we spoke of in Chapter II, Section 1, in the hierarchical case. Indeed, the need will be greater, for larger batteries will be required to reach the same accuracy of estimation with more extra factors.

CHAPTER VIII

MAXIMIZING AND MINIMIZING THE SPECIFICS

1. *A hierarchical battery.*—In Section 3 of Chapter III a brief reference was made to the fact that the Spearman Two-factor Method, and Thurstone's Minimal Rank Method, of factorizing batteries of tests maximize the variance of the specific factors, by reason of minimizing the number of common factors. In the present chapter we shall inquire further into this aspect, and describe a method of estimating factors (Bartlett, 1935, 1937), which in its turn endeavours to minimize the specifics again. First take the case of the analysis of a hierarchical battery. As was illustrated in Chapter III, the analysis of such a battery into one general factor only, and specifics, gives the maximum variance possible to the specifics. The combined communalities of the tests are less in the two-factor analysis than in any other analysis. In the matrix of correlations after it has been reduced to the lowest possible rank, the communalities occupy the principal diagonal:

$$\begin{bmatrix} h_1^2 & r_{12} & r_{13} & r_{14} \\ r_{12} & h_2^2 & r_{23} & r_{24} \\ r_{13} & r_{23} & h_3^2 & r_{34} \\ r_{14} & r_{24} & r_{34} & h_4^2 \end{bmatrix}$$

The mathematical expression of the above fact is that the trace of the reduced correlation matrix, i.e. the sum of the cells of the principal diagonal, is a minimum.

It is true that certain exceptions to this statement are mathematically possible, but their occurrence in actual psychological work is a practical impossibility. They have been investigated by Ledermann (unpublished thesis), who finds, in the case of the hierarchical matrix, that an excep-

tion is only possible when one of the g saturations is greater than the sum of all the others. When the battery is of any size, this is most unlikely to occur: and almost always, when it did occur, the large saturation of one test would turn out to be greater than unity, which is not permissible (the Heywood case).*

2. *Batteries of higher rank.*—The same general statement as the above, that the specifics are maximized, is also true of Thurstone's system, of which its predecessor (Spearman's two-factor system) is a special case. The communalities which give the matrix its lowest rank are in sum less than any other diagonal elements permissible. If numbers smaller than the Thurstone communalities are placed in the diagonal cells, the analysis fails unless factors with a loading of $\sqrt{-1}$ are employed (*Vectors*, page 103), and such factors are, of course, inadmissible.

Here again there are possibly cases where the the lowest rank is not accompanied by the lowest trace (i.e. the lowest sum of the communalities). But here again it seems certain that if such cases do exist, they are mathematical curiosities which would never occur in practice.

As an illustration the reader may use the example of Chapter II, Section 9:

	1	2	3	4	5
1	.	·4	·4	·2	·5883
2	·4	.	·7	·3	·2852
3	·4	·7	.	·3	·2852
4	·2	·3	·3	.	·1480
5	·5883	·2852	·2852	·1480	.

As we there saw, this matrix can be reduced to rank 2 by the unique set of communalities—

·7 ·7 ·7 ·13030 ·5

and we found there that, if we wanted to attain rank 2, we could not, for example, reduce the first communality to ·5.

We can, however, reduce the first communality to ·5 if

* See Chapter XV, Section 5, page 231.

we are willing to accept a higher rank than 2, that is, if we are willing to accept more common factors than two. But we find in that case that the remaining communalities necessarily rise so as to annul, and more than annul, the saving in communality achieved on the first test. We find ourselves bound to take the second communality more than the former ·7, or inadmissible consequences ensue. We have a certain latitude in its choice, but there is a lower limit somewhere between ·7 and ·8 below which it makes the matrix inadmissible. Let us take ·8 as the second communality (having thus still made a gross saving on the former communalities of ·7 and ·7) and calculate the remaining communalities, now fixed, which give rank 3. We can do this by the same process of pivotal condensation used in Chapter II, Section 9, making this time the matrix consist of nothing but zeros after three condensations (for rank 3) and then working back to the communalities. We find for the five communalities—

·5 ·8 ·65474 ·14592 ·80786

with a sum of 2·90852 for the total communality (or trace) compared with the total of 2·73030 with rank 2. Our attempt to save communality by reducing that of the first test from ·7 to ·5 and letting the rank rise has been foiled. The minimum rank carries with it, in all practically possible cases, the minimum communality and the maximum specific variance. *Minimizing the number of common factors maximizes the specific variance.*

3. *Error specifics.*—That some of the variance of a test will probably be unique to that particular test given on that particular occasion is clear; there will be an error specific. But not all errors in testing will produce unique or specific factors. The errors will include sheer blunders, such as mistakes in recording results; sampling errors due to the particular set of persons tested; and variable chance errors in the performances of the individuals. The first can with care be reduced to infinitesimal proportions. Sampling errors will be discussed in Chapter X, and we will only say here that they will in many or most cases produce not specific but common factors. The variable chance

errors in the performances of the individual may be unique to each test, but often they too will run through several tests, as when a candidate has a slight toothache, or is elated by good news, or disturbed by a street organ—all of which things may affect several tests if they are administered on the same day. The "unreliability" of a test, due to variable chance errors, is caused by factors which are unique not to the test but *to the occasion*. Tests a and b performed to-day, and repeated as Tests a' and b' to-morrow, may have reliabilities less than unity, yet the chance errors of to-day may link a and b, and the chance errors of to-morrow may link a' and b'. Nevertheless, some of the error variance will doubtless be unique, but surely nothing like the amount of specific variance due to the Thurstone principle of minimizing the number of common factors can be due to this.

There remains the true specific of each test. It does not seem unreasonable to suppose that such exist, though it is not easy to imagine them existing before the test is given. The ordinary idea of specific factors would be tricks learned by doing that particular test, as a motor-car or a rifle may have and usually does have idiosyncrasies unknown to the stranger. But it seems questionable whether a method of analysis is justifiable which makes specific factors play so large a part.

4. *Shorthand descriptions.*—It is to be observed that an analysis using the minimal number of common factors, and with maximized specific variance, is capable of reproducing the correlation coefficients exactly by means of these few common factors, and in the case of an artificial example will actually do so; while in the case of an experimental example including errors, it will do so at least as well as any other method. If this is our sole purpose, therefore, the Thurstone type of analysis is best, since it uses fewest factors.

But the few common factors of a Thurstone analysis do not enable us to reproduce the original test scores from which we began, they do not enable us to describe all the powers of our population of persons very well. With the same number of Hotelling's "principal components" as

Thurstone has of common factors we could arrive at a better description of the scores, though a worse one of the correlations. The reader may reply that he does not want factors for the purpose of reproducing either the original scores or the original correlations, for he possesses these already! But what we really mean, and what it is very convenient to have, is a concise shorthand description, and the system we prefer will depend largely on our motives, whether we have a practical end in view or are urged by theoretical curiosity. The chief practical incentive is the hope that factors will somehow enable better vocational and educational predictions to be made. Mathematically, however, as we have seen, this is impossible. If the use of factors turns out to improve vocational advice it will be for some other reason than a mathematical one. For vocational or educational prediction means, mathematically, projecting a point given by n oblique co-ordinate axes called tests on to a vector representing the occupation, whose angles with the tests are known, but which is not in the n-space of the tests. The use of factors merely means referring the point in question to a new set of co-ordinate axes called factors, a procedure which cannot define the point any better and, unless care is taken, may define it worse, nor does the change of axes in any way facilitate the projection on to the occupation vector. Moreover, the task of carrying out prediction with the aid of factors is rendered more difficult by the circumstance that the popular systems use more factors than there are tests, so that the factors themselves have to be estimated. In addition, it is usual to estimate only the common factors, throwing away the maximum amount of variance unique to each test, maximized by insisting on as few common factors as possible. If there is any guarantee that these abandoned portions of the test variance are uncorrelated with the occupation to be predicted, no harm is done. But the circumstances under which this guarantee can be given are precisely those circumstances under which a direct prediction without the intervention of factors can easily be made.

5. *Bartlett's estimates of common factors.*—Since, then,

the Thurstone system suffers, from a practical point of view, from this handicap of throwing away all information which can possibly be ascribed, rightly or wrongly, to specific factors, there is a peculiar interest in the proposal (M. S. Bartlett, 1935, 1937*a*, 1938) to estimate the common factors, not by the regression method of the previous chapter, but by a method * which *minimizes* the sum of the squares of a man's specific factors (already *maximized* by the principle of using few common factors).

The way in which Bartlett's estimates differ from regression estimates of factors can be very clearly seen by thinking in terms of the geometrical picture already used in earlier chapters (see Figures 14 to 20). When the factors outnumber the tests, the vectors representing the former are in a space of higher dimensions than the test space.

The individual person is represented in the test space by a point, namely that point P whose projections on to the test vectors give his test scores. We do not know a representative *point* for this individual in the complete factor space, however. His representative point Q may be, for all we know, *anywhere* in the subspace which is perpendicular to the test space and intersects with it at P. In these circumstances the regression method takes refuge in the assumption that this individual is average in all qualities of which we know *nothing*; that is, in all qualities orthogonal to our test space. It therefore assumes P to be his point also in the factor space, and projects P on to the factor axes to get the factor estimates for him.

Bartlett's method is, in the present writer's opinion, equivalent to a different assumption about the position of the point Q. Within the complete factor space there is a subspace which contains the *common* factors. Of all the positions open to the point Q, Bartlett's method chooses that one which is nearest to the common-factor space, and from thence projects on to the common-factor vectors. This is equivalent to making the assumption that this man is *not* average in the qualities about which we know nothing,

* See Appendix, paragraph 18.

but instead possesses in those unknown qualities just those degrees of excellence which bring his representative point to the chosen point Q.

Both the regression method and Bartlett's method make assumptions about qualities which are quite unknown to us, and are quite uncorrelated with the tests we know. The regression assumption is that the man is average in these, Bartlett's assumption is that he is not average ; and because men are most frequently near the average, the regression assumption seems more likely to be correct. The other assumption can be justified only by its utility in attaining special ends ; it cannot be the most generally useful assumption.

6. *Their geometrical interpretation.*—All this can be most clearly seen (because a perspective diagram can be made) in the case of estimating one general factor g only, the hierarchical case. A figure like Figure 19 will illustrate this case, if we take y and z there to be two tests and x to be the g vector (see Figure 21).

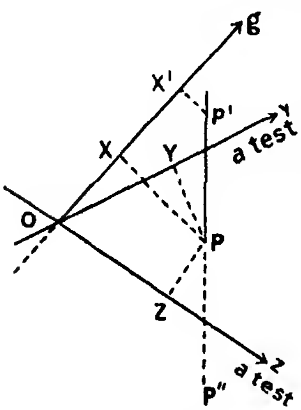

Figure 21.

The man's representative point in the yz plane is P. But we do not know his representative point Q in solid three-dimensional space, only that it is somewhere on the line $P'PP''$. The regression method assumes that it is actually at P, the average, and projects P itself on to the g line to get the estimate OX of g. Bartlett's method, on the other hand, assumes that Q is at that point on $P'PP''$ where it most nearly approaches the g line, that is, somewhere near the position P' in our diagram. Bartlett's estimate of g is then represented by OX'.

Now, any point on the line $P'PP''$, when projected on to the test vectors y and z, gives the same two test scores Y and Z. There is, in general, no point on the line g which does this exactly. But clearly X', of all the points on g,

will be the point whose projections most nearly fall on Y and Z, for X' is as near as possible to the line $P'PP''$. That is, the projection of X' on to the plane of the tests falls as near to the point P as is possible. In other words, if we ignore the specifics entirely and use only the estimated g in the specification of y and z, Bartlett's estimate comes as near as is possible to giving us back the full scores OY and OZ. If the regression estimate OX is projected on to the lines y and z, it will obviously give a worse approximation—much worse in our figure—to OY and OZ.

The regression method, in order to recover as much as possible of the original scores, would have to make a second estimate of them. For the estimates of g represented by quantities like OX are not in standard measure. Before projecting the point X on to the lines y and z, therefore, to recover the original scores as far as possible, the regression method would alter the scale of its space along the g vector until the quantities like OX *were* in standard measure. This would not only change the position of X on the line, it would change the angles which the lines in the figure make with one another; *and would change them exactly in such a manner that, in the new space, the projection of* OX *on to* y *and* z *would fall exactly where the Bartlett projections from* X' *fall in the present space* (Thomson, 1988*a*).

There is, therefore, no final difference in excellence between the two methods in the matter of restoring the original scores as fully as possible, but the regression method takes two bites at the cherry. On the other hand, the regression estimates can be put straight into the specification equation of an occupation which is known to require just these common factors, whereas here it is the Bartlett method which has to have a second shot.

Both methods have to change their estimate of g when a new test is added to the battery. For the man is not very likely to have, in the specific of this new test, either the average value previously assumed by the regression method, or the special value assumed by the Bartlett method. But he is more likely to have the former than the latter, so the Bartlett estimates will change more

than do the regression estimates as the battery grows. Ultimately, when the number of tests becomes infinite, the two forms of estimate will agree.

7. *A numerical example.*—In the case of estimates of one general factor g from a hierarchical battery, the Bartlett estimates differ from the regression estimates only in scale. They put the candidates in the same order of merit for g as do the regression estimates, but give them a greater scatter, making the high g's higher and the low g's lower. The formula is—

$$\frac{1}{S} \Sigma \frac{r_{ig} z_i}{1 - r_{ig}^2}$$

instead of Spearman's—

$$\frac{1}{1 + S} \Sigma \frac{r_{ig} z_i}{1 - r_{ig}^2} \text{ (see page 106).}$$

With more than one common factor, the connexion between the two kinds of estimate is not so simple (Appendix, Section 13). The mathematical reader will be able to calculate the Bartlett factor estimates from the matrix formulæ given in the Appendix. We shall here calculate them, for the example of Chapter VII, Section 5, from the regression estimates there given, and their matrix of variances and covariances given in Section 8 of that chapter.

For if the matrix of regression loadings be represented by N, and the matrix of variances and covariances of the regression factors by K, then the matrix of Bartlett loadings can be shown (Bartlett, 1938) to be—

$$K^{-1}N$$

This matrix multiplication can be carried out by Aitken's pivotal condensation also. For it has been shown (Aitken, 1937*a*) that the pivotal condensation of a pattern of three matrices arranged thus:

$$\begin{array}{c|c} Y & -Z \\ X & . \end{array}$$

gives, when by repeated condensations all numbers have been removed from the left-hand block, the triple product

$X\ Y^{-1}\ Z$. We shall therefore obtain the Bartlett loadings for estimating the factors from the tests if we condense—

$$\begin{array}{c|c} K & -N \\ \hline I & . \end{array}$$

where I is the unit matrix which has unity in each cell of the principal diagonal and zeros elsewhere. The matrices K * and N are taken from pages 125 and 124, and the whole calculation is as follows (to three places of decimals only, to facilitate the arithmetic for readers who wish to check it):

							Check Column	
·674	·218	·127	−·300	−·095	−·532	−·095	−·003	
1·000	·323	·188	−·445	−·141	−·789	−·141	−·004	(5)
·218	·567	−·035	−·353	−·581	·352	·153	·321	
·127	−·035	·556	−·121	·148	·206	−·747	·134	
1·000	.	.	.	.	.	.	1·000	
.	1·000	.	.	.	.	.	1·000	
.	.	1·000	.	.	.	.	1·000	
	·497	−·076	−·256	−·550	·524	·184	·322	(3)
	1·000	−·153	−·515	−1·107	1·054	·370	·650	(49)
	−·076	·532	−·064	·166	·306	−·729	·135	
	−·323	−·188	·445	·141	·789	·141	1·005	
	1·000	.	.	.	.	.	1·000	
	.	1·000	.	.	.	.	1·000	
		·520	−·103	·082	·386	−·701	·184	
		1·000	−·198	·158	·742	−1·348	·354	
		−·237	·279	−·217	1·129	·261	1·215	
		·153	·515	1·107	−1·054	−·370	·351	
		1·000	.	.	.	.	1·000	
			·282	−·180	1·305	−·058	1·299	
			·545	1·083	−1·168	−·164	·297	(6)
			·198	−·158	−·742	1·348	·646	

The Bartlett estimates of the factors, therefore, which

* Slightly corrected to make it symmetrical.

we shall distinguish from the regression estimates by turning the circumflex accent upside down, are—

$$\check{g} = \cdot 232z_1 - \cdot 180z_2 + 1\cdot 305z_3 - \cdot 058z_4$$
$$\check{v} = \cdot 545z_1 + 1\cdot 083z_2 - 1\cdot 168z_3 - \cdot 164z_4$$
$$\check{F} = \cdot 198z_1 - \cdot 158z_2 - \cdot 742z_3 + 1\cdot 348z_4$$

In Chapter VII, Section 5, we imagined a man whose scores in the four tests, in standard deviations, were—

z_1	z_2	z_3	z_4
·2	− ·4	·7	·6

and calculated the regression estimates of his three factors, g, v, and F. By inserting his test scores in the above equations we can find, for comparison, the Bartlett estimates of his factors, shown in the following table:

Factors	g	v	F
Regression estimates	·451	− ·500	·387
Bartlett estimates	·997	− 1·240	·392

This illustrates the tendency of the Bartlett estimates to be farther from the average than the regression estimates are.

PART III

THE INFLUENCE OF SAMPLING AND SELECTION OF THE PERSONS

CHAPTER IX

SAMPLING ERROR AND THE THEORY OF TWO FACTORS

1. *Sampling the population of persons.*—In the previous pages we have seldom mentioned sampling errors. There is an implicit reference to them in Chapter I, where a portion of an actual experimental matrix of correlations is shown as a contrast to the artificial ones used in the text; and later in that chapter there is a closer approach to the difficulties caused by sampling errors. But apart from this, and perhaps one or two other references, the exposition in Parts I and II is entirely free from any consideration of them. The examples are made and worked as if on every occasion the *whole population* of people concerned had been accurately tested.

The advantage of this is that it makes the theoretical principles stand out clearly, unobscured by the sampling difficulty. As a result, to mention one important point, it is thus made clear that the difficulties of estimating factors, described in Chapters VII and VIII, have nothing directly to do with sampling the population, but are due to having more factors than tests. It is true that an absolutely clean cut between an exposition which considers sampling errors, and one which disregards them, cannot be made. For sampling errors introduce error factors, and thereby swell the total number of factors. But even were the whole population of persons tested, factors which outnumber the tests would remain "indeterminate," as it is sometimes expressed, meaning that they can only be estimated, not measured exactly.

Another kind of sampling, however, does exist in Parts I and II, a sampling of the *tests*. We have there assumed that the whole population of persons is tested, but we have not supposed that they were plied with the whole population of tests. It is difficult perhaps to say what

"the whole population of tests" means, but at any rate it is clear that in Parts I and II we were using only a few, not all possible tests. There is thus in our subject a double sampling problem, and this makes it very difficult. In the present section of this book (Part III) we shall consider the effects of sampling the population of *persons*.

The general idea underlying the notion of a sampling error is not a difficult one. Take, for example, the average height of all living Englishmen who are of full age. This could, if need be, be ascertained by the process of measuring every living Englishman of full age. Actually this has never been done, and when anyone makes a statement such as "The average height of Englishmen is 67½ inches," he is basing it upon a sample only. This sample may not be an unbiased one. Indeed, samples of Englishmen whose height has been officially recorded are heavily loaded with certain classes of Englishmen—for example, prisoners in gaol, and unemployed and possibly underfed young men joining the army. The average height of such men may well differ from that of all Englishmen. But when we speak of sampling error, we do not mean error due to the sample being *known* to be a biased one. Even if the sample of Englishmen used to find the average height of their race were, as far as could be seen, a perfectly fair sample, containing the proper proportion of all classes of the community and of all adult ages, etc., it yet would not necessarily yield an average exactly equal to that of *all* Englishmen. Several apparent replicas of the sample would yield different averages. It is these differences, between statistics gathered from different but equally good samples, that we mean by sampling errors.

It is worth while calling attention at this point to a general fact which will be found of importance at a later stage of this book. The *true* average height of Englishmen is only so by definition, and does not in principle differ from the average of a sample. We had to define the population we had in mind as "all living Englishmen of full age." This is a perfectly well-marked body of men. But it is itself in its turn only a sample: a sample of all living Europeans, or all living men. It is, indeed, altering daily

and hourly as men die or reach the age of 21, and each generation is a sample of those that have been and may be. Those who reach the age of 21 are only some, and therefore only a sample, of those born. And even those born are only a sample of those who might have been born had times been better or had there been no war, or a tax on bachelors. So the idea of sampling is a relative one, and the "complete population" from which we take samples is a matter of definition only. The mathematical problem in connexion with sampling which it is desirable to solve if possible for each statistic is to find the complete law of its distribution when it is derived from each of a large number of samples of a given size. Mathematically this is often very difficult, and frequently we have to be content with a formula which gives its approximate variance if certain assumptions are allowed and certain small quantities are neglected.

Sampling problems are of two kinds, direct and inverse. The easier kind of problem is to say what the distribution of a statistic will be in samples of a given size when we know all about the true values in the whole population: the more difficult kind is to estimate what the true value of a statistic is in a complete population when we know its observed value in certain samples. They differ as do problems of interpolation and extrapolation. As an example of the direct kind of problem let us suppose that we actually knew the height of every adult Englishman of full age. We could then, on being told a certain sample of p Englishmen averaged such and such a height, calculate the probability that this sample was a random sample, a probability that would obviously grow less as the average of the sample departed from the average of the whole population. It would also depend on the size of the sample, for if a very large sample deviates far from the true average, it is less likely to be random, more likely to have some reason for the difference, than a small sample with the same average would have.

2. *The normal curve.*—By the distribution of a certain variable in the population we mean the curve (usually expressed as an equation) showing its frequency of occur-

rence for each possible value. Thus the curve in Figure 22 might show the distribution of height in living adult Englishmen, by its height above the base line at each point. More men (represented by the line MN) have the *average* height, 67½ inches, than have the height 73 inches, the frequency of the latter being shown by the line PQ. The shaded area represents all men whose height is 73 inches or more, and its ratio to the area under the whole curve is the probability that an Englishman taken absolutely at random will have a height of 73 inches or more.

Very often distributions are, at any rate approximately, of a certain shape called the "normal curve." The normal curve has a known equation, it is symmetrical about its mid point, and with the aid of published tables can be drawn accurately (or reproduced arithmetically) if we know the mid point M (which is the average of the measurements) and a certain distance ST or $S'T$ (which is equal to the standard deviation of the measurements).

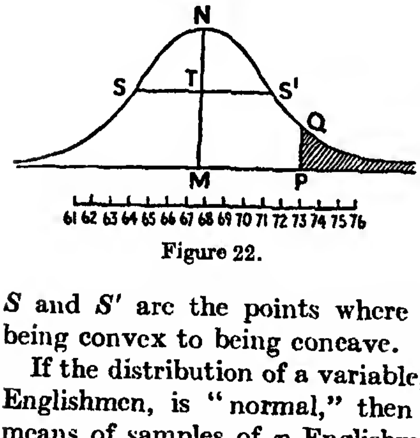

Figure 22.

S and S' are the points where the curve changes from being convex to being concave.

If the distribution of a variable, say the heights of adult Englishmen, is "normal," then the distribution of the means of samples of p Englishmen's heights will also be normal, but will be more closely concentrated about the point M than are the measurements of individuals: in point of fact, its variance will be p times smaller, its standard deviation thus $\sqrt{p}$ times smaller. That is to say, if we take sample after sample of 25 Englishmen each time, and for each sample record the average height, the means thus accumulated will be distributed in a curve of the same shape as that of Figure 22, but narrower from side to side, so that SS' would be one-fifth ($\sqrt{25}$) of what it is in Figure 22, which is the distribution of single measurements.

If a sample were made with some special end in view, such as ascertaining whether red-headed men tend to be tall, we would decide whether we had detected such a tendency by calculating the probability that a mean such as our red-headed sample showed, or a mean still further away from M, would occur at random. For this purpose we would compare the deviation of our sample from M with the standard deviation of the distribution of such samples, obtained by dividing the standard deviation of individuals by the square root of p, the number in the sample. The ratio of the deviation found, to the standard deviation, is the criterion, and the larger it is the more likely is it that red-headed men really do tend to be tall. For most practical purposes we take a deviation of over three times the standard deviation as " significant."

Sometimes the reader will find significance questions discussed in terms of the " probable error " instead of the standard deviation. The probable error is best considered as a conventional reduction of the standard deviation (or standard error, as it is sometimes called) to two-thirds of its value (more exactly, to ·67449 of its value).

Not only would the average height, or the average weight, of the sample of red-headed men differ from sample to sample. Statistics calculated in more complex ways from the measurements will also vary from sample to sample, as, for example, the variance of height, or the variance of weight, or the correlation of height and weight. Let us consider first the variance of the heights. In the whole population this is calculated by finding the mean, expressing every height as a plus or minus deviation from the mean, squaring all these deviations, and dividing the sum by the number in the population.

This is also how we would find the variance of the sample if we really want the variance *of the sample*. But if we want an estimate of the variance in the whole population, and the sample is small, it is better to divide by one less than the number in the sample. A glimpse of the reason for this can be got by considering the case of the smallest possible sample, namely, one man. Here the mean of the sample is the one height that we have measured, and

the deviation of that measurement from the mean of the sample is zero. The formula if we divide by the number in the sample (one) will give zero for the variance—and that is correct for the sample. But it would be too bold to estimate the variance of the whole population from one measurement: if we divide by one less than the sample we get—

$$\text{variance} = \frac{0}{0}$$

that is, we don't know, which is a wiser statement.

The standard error of a variance v, if the parent population from which the samples are drawn is normally distributed, is—

$$\frac{v\sqrt{2}}{\sqrt{p}}$$

where p is the number of persons in the sample. The standard error of a correlation coefficient r is, with the same condition, equal to—

$$\frac{1 - r^2}{\sqrt{p}}$$

In both cases $p - 1$ had better be substituted for p when the samples are small, as a cautionary measure. It is better to magnify than to belittle the errors of our calculations.*

* It is important to remember that sampling the population is not the only source of error in the measurement of statistics, e.g. the correlation coefficient. All sorts of influences may disturb it. These will *usually* "attenuate" the correlation coefficient, i.e. tend to bring it nearer to zero, as can be seen when we consider that a perfect correlation only can be reduced by error. But they will not always do so, and if the errors in the two trait measurements are themselves correlated, they may even increase the true correlations in a majority of cases. An estimate of the amount of variable error present can be made from the correlation of two measurements of the same trait on the same group, a correlation called the "reliability," which should be perfect if no variable errors are present. Spearman's correction for attenuation (see Brown and Thomson, 1925, 156) is based upon this. Like all estimates, the correction for attenuation is correct, even if the errors are uncorrelated, only on the average and not in each instance, and it should never be used unless it is small. If it is large, the experiments are "unreliable" and should be improved.

8. *Error of a single tetrad-difference.*—For our discussion of the influence of sampling on the factorial analysis of tests one of the most important quantities to know is the standard error of the tetrad-difference. There has been much debate concerning the proper formula for this. (See Spearman and Holzinger, 1924, 1925, 1929; Pearson and Moul, 1927; Wishart, 1928; Pearson, Jeffery, and Elderton, 1929; Spearman, 1931.) That generally employed is formula (16) in the Appendix to Spearman's *The Abilities of Man*:

Standard error of $r_{13}r_{24} - r_{23}r_{14}$—

$$\frac{2}{\sqrt{N}}\left[r^2(1 - r_{12} - r_{34} + r^2) + (1 - 2r^2)s^2\right]^{\frac{1}{2}} \quad \text{[Spearman and Holzinger's formula (16).]}$$

where N is the number of persons in the sample,*

r is the mean of the four correlation coefficients, and

s^2 is their mean squared deviation (variance) from r.

The probable error is ·6745 times the above. A worked example will be found on page xii of Spearman's Appendix, using (which is all one can do) the *observed* values of the r's.

It will be remembered that in Section 7 of Chapter I we stated Spearman's discovery in the form "tetrad-differences tend to be zero." If tetrad-differences in the whole population, however, were all actually zero, they would not remain exactly zero in samples, and it is only samples that are available to us. We are faced, therefore, with a twofold problem. (*a*) We have to decide, from the size of the tetrad-differences actually found in our sample, whether the sample is compatible with the theory that the tetrad-differences are zero in the whole population. But (*b*) we should also go on to consider whether the sample is equally compatible with the opposed hypothesis that the tetrad-differences are not zero in the whole population, leaving a verdict of "not proven."

It is very necessary to keep in mind both (a) *and* (b)*:* usually only (*a*) has been considered and (*b*) ignored. To

* We use p to mean the number of persons in this book, but are retaining N here and in "formula 16A" below to preserve the usual appearance of these well-known and much-used expressions.

decide whether a given measured tetrad-difference is compatible with the hypothesis that (when measured in the whole population) it is really zero, all we have to do is to calculate its standard error, and compare it with the value of the tetrad-difference. If the latter is less than three times the standard error, or $4\frac{1}{2}$ times the probable error, the chance of its really being zero is not so small as to rule out that possibility. That is all that this comparison tells us. It *may* be a sampling deviation from zero. For example, if a tetrad is ·065 and its probable error is ·055, it is only 1·2 times its probable error. This means that its true value may very well be zero. But it may, of course, equally well be 2 × ·065 or ·130, which is at the same distance from the observed value as zero is. *It is still more likely to be really* ·065. All that the comparison with the probable error has shown is that the observed value ·065 is *compatible* with a true value of zero.

The importance of not losing sight of this becomes clear when we realize that by taking a sufficiently small sample of people we can raise the probable errors as much as we like. Thus if the samples are small, the observed tetrad-differences are sure to be compatible with the value zero, for their probable errors will be so large. This consideration makes it clear that it is wrong to stop here, as most experimenters have unfortunately done. We must go on to consider (*b*) whether the sample is incompatible with the opposed hypothesis that the tetrad-difference is *not* zero.

Here we are faced with the necessity for some *a priori* decision on what we are going to call *not zero*, just as above we had to make a decision that "1,000 to 1 against" would be the limit of our credulity in accepting a hypothesis as possible. The chance of any observation having been derived from *exactly* the point zero is infinitesimal compared with the sum of the chances of its having come from other values. We must take a region round zero which for practical purposes we are willing to accept as zero. If we take ·05 as the discrepancy from zero which we are in practice willing to accept, we are thereby overlooking a quantity which is something like 10 per cent. of the

average of the correlations we are usually dealing with. This is not a very rigorous demand to make, that the tetrad-difference observed should be incompatible with a hypothesis that the true value is greater than ·05, before we will definitely admit the theory that it is really zero.

This means that the tetrad-difference plus three times its standard error (or 4½ times its probable error) must be within the limit ·05. In the case of the example we quoted above, of ·065 with a p.e. of ·055, this condition is obviously not fulfilled. This tetrad, therefore, is quite compatible with the hypothesis that the true value is *not* zero. We have already shown that it is also compatible with the hypothesis that the true value *is* zero. It is compatible with both hypotheses, and proves neither. And so, indeed, are most tetrad-differences in observed tables of correlations commonly described as hierarchical. In most of them the odds are indeed *against* the hypothesis of zero. All that is meant by describing these tetrad-differences as zero (as is commonly done) is that the odds against that hypothesis are not very heavy, and at any rate not 1,000 to 1 against. Commonly the table is claimed as hierarchical until the odds against that hypothesis rise to 1,000 to 1 or thereabouts. The prisoner is deemed to be hierarchical unless he can produce very strong proof (1,000 to 1) that he is not.

In defence of this practice, which Mr. W. G. Emmett has very strikingly exposed (Emmett, 1936), it may perhaps be urged that the simplest explanation, of one common factor only, is being clung to until the facts force a departure from it to a more complex hypothesis. So long as this is clearly understood by the reader, and he is not misled into thinking that the facts *prove* that only one common factor exists, there is no harm done.*

4. *Distribution of a group of tetrad-differences.*—The actual calculation, for every separate tetrad-difference, of its standard error by Spearman and Holzinger's formula (16) is, however, an almost impossibly laborious task. In a table of correlations formed from *n* tests there are

* For a careful and critical examination of tetrad-difference evidence see Garrett and Anastasi, 1932.

$n(n-1)/2$ correlation coefficients, and $n(n-1)(n-2)(n-3)/8$ different (though not independent) tetrad-differences. Any one particular correlation-coefficient is concerned in $(n-2)(n-3)$ different tetrad-differences, and any one test in $(n-1)(n-2)(n-3)/2$ different tetrad-differences. Thus with ten tests there are 630 tetrad-differences, and with twenty tests 14,535 tetrad-differences. In the latter case, any one test is concerned in 2,907. Under these circumstances, it is natural to look for a more wholesale method than that of calculating the standard error of each tetrad-difference. The method adopted by Spearman is to form a table of the distribution of the tetrad-differences, and compare this distribution with that of a normal curve centred at zero and with standard deviation given by—

$$\frac{2}{\sqrt{N}}\,[r^2(1-r)^2+(1-R)s^2]^{\frac{1}{2}}$$ [Spearman and Holzinger's formula (16A).]

where N = number of persons in the sample,

r = the mean of all the r's in the whole table,

s = their mean squared deviation from r,

$R = 3r\,.\,\dfrac{n-4}{n-2} - 2r^2\,.\,\dfrac{n-6}{n-2}$, and

n = number of tests.

Numerous examples of the comparison of "histograms" of tetrad-differences with normal curves whose standard deviation is found by (16A) are given in Spearman's *The Abilities of Man*. This method of establishing the hypothesis, that the tetrad-differences are derived by sampling from a population in which they are really zero, is open to the same doubt as was explained in the simpler case of one tetrad-difference. The comparison can prove that the tetrad-differences observed are compatible with that hypothesis. It does not in itself prove that they are compatible with that hypothesis only; and, as Emmett has shown in the article already mentioned, the odds are commonly rather against this.

The usual practice, moreover, is to "purify" the battery of tests until the actual distribution of tetrad-differences agrees with (16A), so that in effect all that is then proved is that a team *can* be arrived at which *can* be described in terms of two factors. This, although a more modest claim than has often been made, and certainly less than is implicitly understood by the average reader, is nevertheless a matter of some importance. Not all teams of tests can be explained by one common factor; but it is not very difficult to find teams which can. There is little doubt in the minds of most workers that a *tendency* towards hierarchical order actually exists among mental tests.

5. *Spearman's saturation formula.*—It will be remembered from Section 4 of Chapter I that the calculation of the g saturation of each test forms an important part of the Spearman process. We saw there that in a hierarchical matrix each correlation is the product of the two g saturations of the tests, for example—

$$r_{34} = r_{3g} \cdot r_{4g}$$

Since this is so, each g saturation can be calculated from the correlations of a test with two others, and their inter-correlation. Thus to find r_{1g} we can take Tests 2 and 3 as reference tests, when we have—

$$\frac{r_{12}\, r_{13}}{r_{23}} = \frac{r_{1g}\, r_{2g} \cdot r_{1g}\, r_{3g}}{r_{2g} \cdot r_{3g}} = r_{1g}{}^2$$

When the matrix is really hierarchical, and there are no sampling errors present, it is immaterial which two tests we associate with Test 1 in order to find its g saturation. We have, in fact, in that case—

$$\frac{r_{12} \cdot r_{13}}{r_{23}} = \frac{r_{14} \cdot r_{15}}{r_{45}} = \frac{r_{12} \cdot r_{15}}{r_{25}} = etc.$$

But even if the correlations, measured in the whole population, were really exactly hierarchical, sampling errors would make these fractions differ somewhat from one another, and we are faced with the problem of deciding which value to accept for the g saturation. The average of all possible fractions like the above would be one very

plausible quantity to take but is laborious to compute. Spearman therefore adopts a fraction—

$$\frac{r_{12}\,.\,r_{13} + r_{14}\,.\,r_{15} + r_{12}\,.\,r_{15} + etc.}{r_{23} \quad + \quad r_{45} \quad + \quad r_{25} \quad + etc.} = r_{1g}{}^2$$

whose numerator is the sum of the numerators, and whose denominator is the sum of the denominators, of the single fractions. This combined fraction he computes in a tabular manner which we will next describe, by the algebraically equivalent formula—

$$r_{1g}{}^2 = \frac{A_1{}^2 - A_1{}'}{T - 2A_1}$$ [Spearman's formula (21), Appendix, *Abilities of Man.*]

The quantities A_1, A_2, etc., are the sums of the rows (or columns) of the matrix of correlations without any entries in the diagonal cells. (The arithmetical example is confined to five tests to economize space):

	1	2	3	4	5	A	A^2
1	.	·50	·34	·33	·24	1·41	1·988
2	50	.	·56	·32	·15	1·53	2·341
3	·34	·56	.	·13	·35	1 38	1 904
4	·33	·32	·13	.	·29	1·07	1·145
5	·24	·15	·35	29	.	1·03	1·061
						T = 6·42	

T is the sum of all the A's, and therefore of all the correlations in the table (where each occurs twice). A new table is now written out, with each coefficient squared, and its rows summed to obtain the quantities A':

	1	2	3	4	5	A'
1	.	·250	·116	·109	·058	·533
2	·250	.	·314	·102	·023	·689
3	·116	·314	.	·017	·123	·570
4	·109	·102	·017	.	·084	·312
5	·058	·023	·123	·084	.	·288

The calculation of all the saturations is then best performed in a tabular manner, thus:

	A^2	A'	$A^2 - A'$	$2A$	$T-2A$	$\dfrac{A^2-A'}{T-2A}$	g Saturation
1	1·988	·533	1·455	2·82	3·60	·4042	·66
2	2·341	·689	1·652	3·06	3·36	·4917	·70
3	1·904	·570	1·334	2·76	3·66	·3645	·60
4	1·145	·312	·833	2·14	4·28	·1946	·44
5	1·061	·288	·773	2·06	4·36	·1773	·42

where the last column is the square root of the preceding. The reader should calculate the six slightly different values of r_{1g} from the original table by the formula $(r_{ij} \cdot r_{ik}/r_{jk})^{\frac{1}{2}}$, for comparison with the value ·66 obtained above. He will find—

·55	·72	·89
	·93	·48
		·52

with an average of ·68.

6. *Residues.*—If the correlations which would arise from these saturations or loadings are calculated, and subtracted from the observed correlations, we obtain the residues which have then to be examined to see if they are small enough to be attributable to sampling error. In the following double table of correlations are set out the observed correlations uppermost, and those calculated from the g saturations below. The difference is the residue, which may be plus or minus :

g Loadings	·66	·70	·60	·44	·42
·66	.	·50	·34	·33	·24
		·46	·40	·29	·28
·70	·50	.	·56	·32	·15
	·46		·42	·31	·29
·60	·34	·56	.	·13	·35
	·40	·42		·26	·25
·44	·33	·32	·13	.	·29
	·29	·31	·26		·18
·42	·24	·15	·35	·29	.
	·28	·29	·25	·18	

The lower numbers are the products of the two saturations. In this case the residues range from $-\cdot 14$ to $+\cdot 14$ and at first sight appear in many cases to be too large to be neglected in comparison with the original correlations. To check this impression, the standard errors of the latter have to be calculated by the formula—

$$\frac{1 - r^2}{\sqrt{p}}$$

where p is the number of persons tested, here 50. The standard error of ·56 is therefore ·10, and the residue ·14 is well within three times the standard error. But as the reader will observe, this conclusion is due more to the large size of the standard error than to the small size of the residue. The residue is here *attributable* to sampling error, because the latter is so large. But because the latter is large it does not follow that the large residue is certainly due to it. A test of the second kind is needed here (but is hardly ever applied) to determine the odds for or against the alternative hypothesis, that the residue is *not* due to sampling error. The lack of tests of this second kind, as has already been emphasized in discussing tetrad-differences, is one of the most serious blemishes in the treatment of data during factorial analysis. If we are willing to allow 10 per cent. of the correlation coefficient as being a negligible quantity (a very generous concession), then the chance of our experimental value ·56 having come by sampling from *outside* the area $\cdot 42 \pm \cdot 042$ is (with 50 cases in the sample) still quite considerable, about 5 to 1 *for*. These odds do not justify us in feeling confident that ·56 *does* come from outside $\cdot 42 \pm \cdot 042$. But much less do they justify us in feeling that it comes from inside that region.

7. *Reference values for detecting specific correlation.*—If, after a calculation like that described, one of the residues is found to be too large to be explicable by sampling error, the excess of correlation over that due to g is attributed to "specific correlation," meaning correlation due to a part of their specific factors being not really unique but shared by these two tests. In the case of our numerical example,

if the number of subjects tested had been larger, the standard errors of the coefficients would have been smaller, and some of the discrepancies between the experimental values and those calculated from the g saturations would have been too large to be overlooked, but would have had to be attributed to specific correlation. In such a case, the g loadings would, of course, be wrong and would have to be recalculated from the battery after one of the tests concerned in the specific correlation was removed from it. Later, the other test could be replaced in the battery instead of the first, and thus *its* g saturation found. The difference between the experimental correlation of the two, and the product of their g saturations, with a standard error dependent on the size of the sample, would be then attributed to their specific linkage.

If two tests, v and w, are thus suspected of having a specific link as well as that due to g, it is clear that the smallest battery of tests which could be used in the above manner to detect that link would be one of *two* other tests, x and y, say, to make up a tetrad:

	v	x
w	r_{vw}	r_{xw}
y	r_{vy}	r_{xy}

and these two "reference" tests would have to be known to have no specific links with each other or with the two suspected tests. The example which gave rise to Figure 5 (see Chapter I, page 15) illustrates this. Tests 2 and 3 there are, let us suppose, those with a suspected specific link. The tetrad-difference to be examined by means of Spearman's formula (16) is that which has r_{23} as one corner. In such a case, where the two reference tests 1 and 4 are known to have no link except g with one another, or with the other two tests, two of the possible tetrad-differences ought to be larger than three times the standard error given by formula (16), and equal to one another, while the third tetrad-difference should be zero (or sufficiently near to zero, in practice) (Kelley, 1928, 67).

The g saturation of each of the tests under examination

for specific correlation can be found by grouping it with the two reference tests. Thus in the case of our Figure 5, we have—

$$r_{2g}{}^2 = \frac{r_{12} \cdot r_{24}}{r_{14}} = \frac{\cdot 5 \times \cdot 5}{\cdot 5} = \cdot 5$$

$$r_{3g}{}^2 = \frac{r_{13} \cdot r_{34}}{r_{14}} = \frac{\cdot 5 \times \cdot 5}{\cdot 5} = \cdot 5$$

Therefore the correlation between 2 and 3 which is due to g is—

$$r_{2g} \cdot r_{3g} = \sqrt{\cdot 5} \times \sqrt{\cdot 5} = \cdot 5$$

and the difference between this and ·8, the actual value, is the part to be explained by the specific factor shared by these two tests. The difference of ·3 is not what is called the specific correlation itself, it should be remarked, but only its numerator. By specific correlation is meant the correlation between the two "specific" parts of the linked tests, due to these not being entirely unique, but having a part in common. How to calculate this we shall see after considering the effect of selection on correlation, in Chapter XI, end of Section 2 (page 173).

When there are several reference tests available, all believed to have no link except g with one another or with the two tests suspected of specific overlap, there will be a number of ways of picking two of them to obtain the tetrad required to decide the matter, and the results will, because of sampling and other errors, be discrepant. Under these circumstances Spearman has devised an interesting procedure for amalgamating the results into one, which we can describe with the aid of the Pooling Square. Instead of using two single tests, let us in the first place imagine that the n tests available as reference tests are divided into two pools equal in number $\left(\frac{n}{2}\right)$, and that the correlations of these pools with one another, and with the suspected tests, are used to form the tetrad. Following Spearman's notation in paragraph 9 of the Appendix to *The Abilities of Man*, we shall call the suspected tests v and w, and the

two pools the x pool and the y pool. We then want the tetrad of correlation coefficients :

	v	x *pool*
w	r_{vw}	r_{xw}
y *pool*	r_{vy}	r_{xy}

of which r_{vw} is known experimentally. The others we can find by using pooling squares. Take first r_{xy}. We have (writing three tests in each reference pool instead of $\frac{n}{2}$) :

$$\begin{array}{c|ccc|ccc}
 & x_1 & x_2 & x_3 & y_4 & y_5 & y_6 \\
\hline
x_1 & 1 & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} \\
x_2 & r_{12} & 1 & r_{23} & r_{24} & r_{25} & r_{26} \\
x_3 & r_{13} & r_{23} & 1 & r_{34} & r_{35} & r_{36} \\
\hline
y_4 & r_{14} & r_{24} & r_{34} & 1 & r_{45} & r_{46} \\
y_5 & r_{15} & r_{25} & r_{35} & r_{45} & 1 & r_{56} \\
y_6 & r_{16} & r_{26} & r_{36} & r_{46} & r_{56} & 1
\end{array} .$$

and the correlation r_{xy} of the two pools with one another is (Chapter VI, Section 2)—

$$r_{xy} = \frac{\left(\frac{n}{2}\right)^2 \bar{r}_c}{\left[\frac{n}{2} + \left\{\left(\frac{n}{2}\right)^2 - \frac{n}{2}\right\} \bar{r}_a\right]^{\frac{1}{2}} \left[\frac{n}{2} + \left\{\left(\frac{n}{2}\right)^2 - \frac{n}{2}\right\} \bar{r}_b\right]^{\frac{1}{2}}}$$

Here the quantities $\bar{r}_a$, $\bar{r}_b$, and $\bar{r}_c$ are the *mean* values of the correlation coefficients (excluding the units) to be found in the quadrants of the pooling square, thus :

$$\begin{array}{c|c}
\bar{r}_a & \bar{r}_c \\
\bar{r}_c & \bar{r}_b
\end{array}$$

Now, there is clearly an arbitrary factor left in this procedure, inasmuch as the division of the n available tests into an x pool and a y pool can be made in many different ways, in each of which the mean values $\bar{r}_a$, $\bar{r}_b$, and $\bar{r}_c$ will be slightly different. To obviate this, Spearman takes the mean value $\bar{r}$ of *all* the n reference tests with one another

instead of each of these three means, upon which the formula for r_{xy} simplifies to—

$$r_{xy} = \frac{\frac{n}{2}\bar{r}}{1 + \left(\frac{n}{2} - 1\right)\bar{r}}$$

and it is this value which he uses in the tetrad.

Similarly, the correlation of the test w with the x pool can be found by a pooling square:

	w	x_1	x_2	x_3
w	1	r_{w1}	r_{w2}	r_{w3}
x_1	r_{w1}	1	r_{12}	r_{13}
x_2	r_{w2}	r_{12}	1	r_{23}
x_3	r_{w3}	r_{13}	r_{23}	1

Its value is—

$$r_{xw} = \frac{\frac{n}{2}\bar{r}_w}{\left[\frac{n}{2} + \left\{\left(\frac{n}{2}\right)^2 - \frac{n}{2}\right\}\bar{r}\right]^{\frac{1}{2}}} = \frac{\left(\frac{n}{2}\right)^{\frac{1}{2}}\bar{r}_w}{\left[1 + \left(\frac{n}{2} - 1\right)\bar{r}\right]^{\frac{1}{2}}}$$

Here, for the same reason as before, not only do we use the average inter-correlation of all the reference tests for $\bar{r}$, but for $\bar{r}_w$ we use the average of the correlations of the test w with *all* the reference tests and not merely with the x pool, for the x pool could be *any* half of them.

Similarly the correlation r_{vy} is found. Thus to form the tetrad all that we need do is to find:

$\bar{r}$, the average correlation of all the reference tests with one another;

$\bar{r}_w$, the average correlation of all the reference tests with w;

$\bar{r}_v$, the average correlation of all the reference tests with v;

and substitute in the formulæ. A numerical example is given by Spearman on page xxii of his Appendix.

CHAPTER X

MULTIPLE-FACTOR ANALYSIS WITH FALLIBLE DATA

1. *Method of approximating to the communalities.*—The influence of sampling errors on multiple-factor analysis is in general similar to that on the tetrad method. Sampling errors blur the picture. They make it both difficult for us to see the true outlines and easy to entertain hypotheses which cannot be disproved, though often they cannot be proved either, by the data.

With artificial data like the examples used in Chapter II it may be laborious, but is not impossible, to find the actual rank of the matrix with various communalities, and thus to arrive by trial at the minimum rank. But when sampling errors are present, or any kind of errors, the question becomes at once immensely more difficult. We have seen in the previous chapter something of the difficulty of deciding from the size of the tetrad-differences when the rank of a matrix may justifiably be regarded as *one*. Such methods have not been used for higher ranks. The labour of calculating all three-rowed, four-rowed, or larger minors, setting out their distribution and comparing it with that to be anticipated from true zero values plus sampling error is too great, and the mathematical difficulty not slight. What has been done is to judge of the rank by the inspection of the residues left after the removal of so-and-so many common factors, e.g. at the end of so-and-so many cycles of Thurstone's process, just as in Section 6 of the preceding chapter we examined the residues left after one common factor was removed. But we must first show how Thurstone meets the difficulty of the unknown communalities.

His practice is to use as an approximate communality the largest correlation coefficient in the column (*Vectors*, 89). That this is a plausible approximation can be seen

from the following considerations. If there were only one general factor, the communality of Test 1 would be—

$$\frac{r_{12} \cdot r_{13}}{r_{23}}$$

where Tests 2 and 3 are any two other tests of the battery. If we take those two other tests which have the highest correlation with Test 1, they are rather likely to have a high correlation with one another. In that case r_{12}, r_{13}, and r_{23} will be much of a size, and—

$$\frac{r_{12} \cdot r_{13}}{r_{23}}$$

reduces approximately to either r_{12} or r_{13}, which are the highest correlations in the column.

We shall illustrate this approximate method of Thurstone's on the same example as we used near the end of Chapter II, for the sake of comparison and for ease in arithmetical computation, even although that example is really an exact and artificial one unclouded by sampling error. Inserting then the highest coefficients in each column we get:

	(·5883)	·4	·4	·2	·5883	
	·4	(·7)	·7	·3	·2852	
	·4	·7	(·7)	·3	·2852	
	·2	·3	·3	(·3)	·1480	
	·5883	·2852	·2852	·1480	(·5883)	
	2·1766	2·3852	2·3852	1·2480	1·8950	= 10·0900
						= $3{\cdot}1765^2$
First Loadings	·6852	·7509	·7509	·3929	·5966	

The communalities which really give the minimum rank are, as we saw in Section 9 of Chapter II—

·7 ·7 ·7 $\cdot 130\dot{3}$ ·5

and the correct first-factor loadings obtained by their use—

·7257 ·7564 ·7564 ·3420 ·5729

With a large battery the difference between the loadings obtained by the approximation and by the correct com-

munalities would be much less. For the "centroid" method depends on the *relative* totals of the columns of the correlation matrix; and when there are twenty or more tests, these relative totals will not be seriously changed by the exact value given to the communality in the column. When the number of tests is large, the influence of the one communality in each column is swamped by the influence of the numerous correlations.

The process now goes on as in Chapter II, and the residuals left after subtraction of the first-factor matrix check by summing in each column to zero, as there.

Before, however, proceeding any farther, in this approximate method *we delete the quantities in the diagonal* (the residues of the guessed communalities) *and replace them by the largest coefficient in the column* regardless of its sign, which we change to plus in the diagonal cell if it is negative in its own cell. The reason for this is apparent, especially when, as may and does happen, the existing diagonal residues are negative, which is theoretically impossible. For although the guessing of the first communalities does not in a large battery make much difference to the first-factor loadings, it may make a big difference to the diagonal residues. If the battery is very large indeed, our first-factor loadings would come out much the same, even if we entered *zero* for every communality, but the diagonal residues would then all be negative. In short, the diagonal residues are much the least trustworthy part of the calculation when approximate communalities are used, and it is better to delete them at each stage and make a new approximation.

2. *Illustrated on the Chapter II example.*—To make this clearer, the whole approximate process is here set out for our small example as far as the second residual matrix. The explanations printed alongside the calculation will make each stage clear. It is important to form the residual matrices exactly as instructed, as otherwise the check of the columns summing to zero will not work. In practice, certainly if a calculating machine were being used, several of the matrices here printed for clearness would be omitted; for example, with a machine one would go straight from

A to *C*, while *D* and *E* would be made by actually altering *C* itself:

A		(·5883)	·4	·4	·2	·5883	Largest *r* of column inserted in diagonal cell.
		·4	(·7)	·7	·3	·2852	
		·4	·7	(·7)	·3	·2852	
		·2	·3	·3	(·3)	·1480	
		·5883	·2852	·2852	·1480	(·5883)	
		2·1766	2·3852	2·3852	1·2480	1·8950	= 10·0900
							= 3·1765²
Loadings I		·6852	·7509	·7509	·3929	·5966	= 3·1765
B	·6852	(·4695)	·5145	·5145	·2692	·4088	First-factor matrix.
	·7509	·5145	(·5639)	·5639	·2950	·4480	
	·7509	·5145	·5639	(·5639)	·2950	·4480	
	·3929	·2692	·2950	·2950	(·1544)	·2344	
	·5966	·4088	·4480	·4480	·2344	(·3559)	
C		(·1188)	−·1145	−·1145	−·0692	·1795	First residual matrix. *A* − *B*
		−·1145	(·1361)	·1361	·0050	−·1628	
		−·1145	·1361	(·1361)	·0050	−·1628	
		−·0692	·0050	·0050	(·1456)	−·0864	
		·1795	−·1628	−·1628	−·0864	(·2324)	
		·0001	−·0001	−·0001	·0000	−·0001	Columns check to zero.
D		(·1795)	−·1145	−·1145	−·0692	·1795	Largest *r* of each column (regardless of sign) inserted in each diagonal cell.
		−·1145	(·1628)	·1361	·0050	−·1628	
		−·1145	·1361	(·1628)	·0050	−·1628	
		−·0692	·0050	·0050	(·0864)	−·0864	
		·1795	−·1628	−·1628	−·0864	(·1795)	
		·6572	·5812	·5812	·2520	·7710	Sum disregarding signs.
E		(·1795)	·1145	·1145	·0692	·1795	Signs of Tests 2, 3, and 4 changed to make largest column (·7710) all positive.
		·1145	(·1628)	·1361	·0050	·1628	
		·1145	·1361	(·1628)	·0050	·1628	
		·0692	·0050	·0050	(·0864)	·0864	
		·1795	·1628	·1628	·0864	(·1795)	
Algebraic Sum		·6572	·5812	·5812	·2520	·7710	= 2·8426
							= 1·6860²
Loadings II		·3898	·3447	·3447	·1495	·4578	(With temporary signs.)

F	·3898	(·1519)	·1844	·1844	·0583	·1783	Second-factor matrix, using temporary signs.
	·3447	·1844	(·1188)	·1188	·0515	·1576	
	·3447	·1844	·1188	(·1188)	·0515	·1576	
	·1495	·0583	·0515	·0515	(·0124)	·0688	
	·4578	·1783	·1576	·1576	·0688	(·2091)	
G		(·0276)	—·0199	—·0199	·0109	·0012	Second residual matrix. $E - F$
		—·0199	(·0440)	·0173	—·0465	·0052	
		—·0199	·0173	(·0440)	—·0465	·0052	
		·0109	—·0465	—·0465	(·0640)	·0180	
		·0012	·0052	·0052	·0180	(—·0296)	
		—·0001	—·0001	·0001	—·0001	·0000	Columns check to zero.

Notes.—It is fortuitous that *all* the entries in E are positive. Usually some will be negative.

In the check for the residual matrices, a discrepancy from zero in the last figure is often to be expected, even of three or four units in a large matrix.

Note the negative value occurring in a diagonal cell in G.

Further stages would be carried on in the same way. But at each stage the residues will be examined, *in comparison with the standard errors of the original correlation coefficients,* to see if further analysis is worth while. Let us do so with the residues of matrix G.

For this purpose let us assume that our experimental correlations were obtained from a population of 900 persons. The standard errors of the correlation coefficients are to be calculated from the formula—

$$\frac{1 - r^2}{\sqrt{p}} = \frac{1 - r^2}{30}$$

The following table shows three times the standard errors of the original correlation coefficients:

.	·084	·084	·096	·065
·084	.	·051	·091	·092
·084	·051	.	·091	·092
·096	·091	·091	.	·098
·065	·092	·092	·098	.

and it will be seen that all the numbers in the matrix of

second residues G are well below these values. Further analysis would therefore be illusory.

The matrix of loadings of common factors thus arrived at is, after we have replaced the proper signs in Loadings II:

Test	*Approximate Method*			*True Values*
	I	*II*	*Communality*	*Communality*
1	·6852	·3898	·6214	·7000
2	·7509	− ·3447	·6827	·7000
3	·7509	− ·3447	·6827	·7000
4	·3929	− ·1495	·1767	·1303
5	·5966	·4573	·5651	·5000
			2·7286	2·7303

The communalities ·6214, etc., are the sums of the squares of the two loadings. For comparison with the approximate communalities thus obtained there are shown the true values, which in this artificial case are known to us (see Chapter II, Section 9). This is for instructional purposes only—the comparison is not intended as any criticism of Thurstone's method of approximation. As has been explained, this method is used only on large batteries, and it is a very severe test indeed to employ it on a battery of only five tests.

We might now go back and begin our whole calculation again, using the communalities ·6214, etc., arrived at by the first approximation. This does not seem often to be done in practice, most workers being content with the approximation first arrived at. If we repeat the calculation again and again with our present example, on each occasion using as communalities the sum of the squares of the loadings given by the preceding calculation, we get the following sets of closer and closer approximation to the true communalities: *

* It is perhaps worth while noting here a surmise of the present author's, based on trials only and without any rigorous algebraic foundation, that when this process is used on a matrix containing

	h_1^2	h_2^2	h_3^2	h_4^2	h_5^2
First trial communalities	·5883	·7000	·7000	·3000	·5883
Next approximation	·6214	·6827	·6827	·1767	·5651
Next approximation	·6381	·6970	·6970	·1477	·5392
Next approximation	·6535	·7043	·7043	·1397	·5258
True values	·7000	·7000	·7000	·1303	·5000

The example has served to show how to work Thurstone's method of approximating to the communalities. It should be emphasized again that, being composed of only five tests, it is not a suitable example to employ in criticism of that method, and it is not so used here, but only as an illustration. Being an artificial example, and not really overlaid with sampling error, it has had the advantage of allowing us to compare the approximations with the true values. But it must be remembered that a real experimental matrix is not likely to have an *exact* low rank to which approximation can converge as here. In that case the approximations will presumably give an indication of the low rank which the matrix *nearly* has, which it might be made to have by adjustments in its elements within the limits of their sampling errors.

We might, indeed, have dealt with this method in Chapter II, quite unconnected with sampling errors, regarding it as a method of finding the communalities by successive approximations. It has, however, been left to the present chapter because in actual practice it is associated with the difficulty of finding communalities because of sampling error, and also is not generally used as a repetitive process. The labour of repeating the whole calculation with new approximations to the communalities has been a deterrent, and the further fact that with large batteries the improvement produced is very small. Usually, therefore, the experimenter is content with the factor loadings first obtained. It is a great drawback of the method, especially in this form, that any mathematical

too few tests, so that there are many alternative sets of communalities giving the lowest rank, it converges to that set which gives the minimum trace, i.e. the minimum total communality and maximum total specific variance.

expression of the standard errors of the resulting loadings is almost impossible, by reason of the chance nature of the approximations made at each stage. On the other hand, the method does give loadings which will imitate the experimental correlations to any desired degree of exactness, and does so with not very laborious arithmetic.

3. *Error specifics.*—We shall consider next the influence of sampling errors upon the specific factors of tests. It has already been remarked (Chapters III and VIII) that these factors play an important part in Spearman's and Thurstone's methods of analysis, which make them as large as possible. We have hitherto used the term "specific" in specific factor and specific variance to mean all that part of a test ability which is unique to that test, or even, in Thurstone's system, which can by the utmost constraining be treated as unique to that test. There is a tendency, however, to confine the term *specific* factor to that non-communal part of the test ability which is not due to any kind of error, and to use "uniqueness" for the whole of what we have hitherto called specific, for both the true specific and the error specifics, as we might put it.

In several places in *The Vectors of Mind* Thurstone emphasizes his point that every test is sure to have some unique variance. This uniqueness he analyses (page 78 and page 180) into three parts: "(*a*) The variable chance errors in the scores of the individuals; (*b*) The specific factors or abilities which are almost certain to be involved in each test of any finite battery; and (*c*) The sampling errors in the coefficients of correlation. All three of these sources of variance," he continues, "are unique for each test; and hence they must be accounted for by unique factors, i.e. factors which are, by definition, not common factors."

In this analysis, both (*a*) and (*c*) are in different senses sampling errors. The errors (*a*) arise because on any one occasion we only have a sample of the performance of each individual, and his powers vary from occasion to occasion. It is, however, (*c*) which particularly concerns us in the present chapter, for the part (*c*) of this analysis of the uniqueness is due to sampling errors in the correlation

coefficients, that is, due to only a sample of the population being tested. Now it is not at all obvious that such sampling errors in the correlation coefficients will produce, as Thurstone says, unique factors. Rather the contrary. In general, they will produce new common factors, for the sampling errors of correlation coefficients are themselves correlated. Pearson and Filon gave the formulæ for such correlation in 1898. The correlation coefficient of the sampling errors of r_{12} and r_{13} (where one of the tests occurs in each correlation) is given by—

$$r_{r_{12}r_{13}} = r_{23} - (\text{a complicated function of } r_{12}, r_{13}, r_{23})$$

and is roughly somewhat less than r_{23}, therefore, for positive correlations.* The correlation coefficient of the sampling errors of r_{12} and r_{34}, on the other hand, is a much smaller quantity of the second order only. The result of this is (Thomson, 1919*a*, 406) that in a table of positive correlations like this, where for greater clearness the subscripts have been omitted :

	1	2	3	4	5	6	7
1	.	r	r	r	r	r	r
2	r		r	r	r	r	r
3	r	r	.	r	r	r	r
4	r	r	r	.	r	r	r
5	r	r	r	r		r	r
6	r	r	r	r	r	.	r
7	r	r	r	r	r	r	.

if r_{25}, say, happens to be the coefficient with the *largest* sampling error, then because of the fact expressed by Pearson and Filon's formula, all the coefficients in the row and column which cross at r_{25} will tend to have large sampling errors in the same direction, while the other correlation coefficients will tend to have smaller sampling

* For positive correlations the "complicated function" referred to is non-negative.

errors. The sampling errors thus tend to produce, not irregular ups and downs of the correlations, but a ridged effect, with a general upward, *or* a general downward, tendency. In other words, the error factors are, or include, *common* factors. Some of the unique variance of the tests may be due to sampling errors: but so will some of the communality of the tests. The effect of sampling errors on factors and factorial analyses is indeed a very complex business, and before we consider it further it is advisable to discuss how deliberate selection of the population (whether human selection or natural selection) modifies analyses. We shall do this in Chapter XI, where selection in one trait only is considered, and in Chapter XII, where the more complex question of simultaneous selection in several traits is dealt with.

CHAPTER XI

THE INFLUENCE OF UNIVARIATE SELECTION ON FACTORIAL ANALYSIS *

1. *Univariate selection.*—All workers with intelligence tests know, or ought to know, that the correlations found between tests, or between tests and outside criteria, depend to a very great extent indeed upon the homogeneity or heterogeneity of the sample in which the correlations were measured. If, to take the usual illustration, we measure the correlation between height and weight in a sample of the population which includes babies, children, and grown-ups, we shall obviously get a very high result. If we confine our measurement to young people in their 'teens, we shall usually get a smaller value for the coefficient of correlation. If we make the group more homogeneous still, taking, say, only boys, and all of the same race and exactly the same age, the correlation of height and weight will be still less.† Through all these changes towards greater homogeneity in age, the standard deviation (or its square, the variance) of height has also been sinking, and the standard deviation of weight also. The formulæ which describe these changes (in samples normally distributed, at any rate) were given in 1902 by Professor Karl Pearson, and when the selection of the persons forming the sample is made on the basis of one quality only, these formulæ can be put into the following very simple form.

Let the standard deviations of (say) four qualities be in the complete population—we must, of course, in each case define what we mean by the complete population, as for example all living adults who were born in Scotland—given by Σ_1, Σ_2, Σ_3, and Σ_4, and their correlations by

* Thomson, 1937 and 1938*b*.

† Greater homogeneity need not *necessarily*, in the mathematical sense, decrease correlation, and occasionally it does not do so in actual psychological experiments. But it almost always does so.

R_{13}, R_{14}, etc. Now let a selection of persons be made who are more homogeneous in the first quality—say, in an intelligence test which has been given to them all—so that its standard deviation in the sample is only σ_1, and write—

$$\frac{\sigma_1}{\Sigma_1} = p_1$$

The smaller p_1 is, the more homogeneous the group is in intelligence-test score. If we write—

$$q_1 = \sqrt{(1 - p_1^2)}$$

q_1 will be larger, the greater the shrinkage in intelligence score-scatter from Σ_1 to σ_1. We shall call q_1 the " shrinkage " of the quality No. 1 in the sample.

The other qualities 2, 3, and 4, being correlated with the first, will tend to shrink with it, and their expected shrinkages q_2, q_3, and q_4 can be calculated from the formula—

$$q_i = q_1 R_{1i}$$

For the sort of reason indicated earlier in this paragraph, the correlations of the four qualities—which we are for simplicity in exposition assuming to be *positively* correlated in the whole population—will also alter, according to the formula—

$$r_{ij} = \frac{R_{ij} - q_i q_j}{p_i p_j}$$

A numerical example will illuminate these formulæ. Let us define our " whole population " as all the eleven-year-old children in Massachusetts, and let us suppose (the numbers are entirely fictitious) that the standard deviations of all their scores in four tests are :

1.	Stanford-Binet test	$16{\cdot}5 = \Sigma_1$,
2.	The X reading test	$24{\cdot}9 = \Sigma_2$,
3.	The Y arithmetic test	$27{\cdot}3 = \Sigma_3$,
4.	The Z drawing scale	$14{\cdot}2 = \Sigma_4$,

while the correlations between these four, in a State-wide survey, are (these are the R correlations):

	1	2	3	4
1	.	·69	·75	·32
2	·69	.	·54	·18
3	·75	·54	.	·06
4	·32	·18	·06	.

Now let a sample of Massachusetts eleven-year-olds be taken who are less widely scattered in intelligence, with a standard deviation in their Stanford-Binet scores of only 10·2. How will all the other quantities listed above tend to alter in this sample ? We have, using the formulæ quoted, the following—

$$p_1 = \frac{10 \cdot 2}{16 \cdot 5} = \cdot 618$$

$$q_1 = \sqrt{(1 - \cdot 618^2)} = \cdot 786$$

and from $q_i = q_1 R_{1i}$ we have the other shrinkages q, and thence the coefficients p and the new standard deviations $\sigma = p\Sigma$:

	1	2	3	4
q	·786	·542	·590	·252
p	·618	·840	·808	·968
σ	10·2	20·9	22·1	13·7

The formula for r_{ij} then enables us at once to calculate the correlations to be expected in the sample, namely :

	1	2	3	4
1	.	·521	·574	·204
2	·521	.	·325	·054
3	·574	·325	.	—·113
4	·204	·054	—·113	.

The greater homogeneity in the sample has made all the correlation coefficients smaller, and has indeed made r_{34} become negative.

2. *Selection and partial correlation.*—If a sample is made completely homogeneous in the Stanford-Binet test, clearly $p_1 = 0$ and $q_1 = 1$. The same formulæ then give us :

	1	2	3	4
q	1	·69	·75	·32
p	0	·524	·488	·904
σ	0	13·0	11·9	12·8

and the resulting correlation coefficients, which in this case are called " coefficients of partial correlation for *constant* Stanford-Binet score," are, by the same formula :

	1	2	3	4
1	.	.	.	.
2	.	.	·098	— ·086
3	.	·098	.	— ·455
4	.	— ·086	— ·455	.

The correlations of the Stanford-Binet test with the others are given by the formula as 0/0, that is, indeterminate. That they are really zero is seen from the fact that when p_1 is taken as not quite zero, but very small, these correlations come out by the formula as very small. They vanish with p_1.

In this special case of " partial correlation," where the directly selected test is so stringently selected that everyone in the sample has exactly the same score in it, our formula—

$$r_{ij} = \frac{R_{ij} - q_i q_j}{p_i p_j}$$

has a more familiar form. For since—

$$q_i = q_1 R_{1i}$$

and $$q_1 = 1$$

in this case of complete shrinkage we have—

$$q_i = R_{1i}$$

and $$p_i = \sqrt{(1 - R_{1i}^2)}$$

so that our formula becomes—

$$r_{ij} = \frac{R_{ij} - R_{1i}R_{1j}}{\sqrt{(1 - R_{1i}^2)}\sqrt{(1 - R_{1j}^2)}}$$

the usual form of a partial correlation coefficient. Its more conventional notation is, calling the test which is made constant test k instead of Test 1—

$$r_{ij \cdot k} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{(1 - r_{ik}^2)}\sqrt{(1 - r_{jk}^2)}}$$

If the " test " which is held constant is the factor g, this becomes—

$$r_{ij \cdot g} = \frac{r_{ij} - r_{ig}r_{jg}}{\sqrt{(1 - r_{ig}^2)}\sqrt{(1 - r_{jg}^2)}}$$

which is called the " specific correlation " between i and j. As we said at the close of Chapter VIII, its numerator is the " residue " left after removing the correlation due to g. If g is the sole cause of correlation, holding g constant will destroy the correlation and we shall have—

$$r_{ij} = r_{ig}r_{jg}$$

as we already saw from another point of view was the case in a hierarchical battery, in Section 4 of Chapter I.

3. *Effect on communalities.*—The formula—

$$r_{ij} = \frac{R_{ij} - q_i q_j}{p_i p_j}$$

is thus a very useful formula, including partial correlation as a special case. If the original variances are each taken as unity, the numerator $R_{ij} - q_i q_j$ for $i \neq j$ gives the new covariances, while p_i^2 and p_j^2 are the new variances.

It also includes as a special case the formula known as the Otis-Kelley formula, which is applicable when two variates have both shrunk to the same extent (a restriction not always recognized). If we put $q_i = q_j$ and therefore $p_i = p_j$ it becomes—

$$p^2 r_{ij} = R_{ij} - q^2 = R_{ij} - 1 + p^2$$

$$p^2(1 - r_{ij}) = 1 - R_{ij}$$

$$\frac{1 - R_{ij}}{1 - r_{ij}} = p^2 = \frac{\sigma_i^2}{\Sigma_i^2} = \frac{\sigma_j^2}{\Sigma_j^2} \text{ the Otis-Kelley formula.}$$

It has a still further application (Thomson, 1938*b*, 456), for if a matrix of correlations in the wider population has been analysed by Thurstone's process, this same formula gives the new communalities (with one exception) to be expected in the sample, if we put $i = j$ and understand by R_{ii}, the communality in the wider population, by r_{ii}, the

communality in the sample (and not a reliability coefficient, which is the usual meaning of this symbol). Writing the usual symbol h^2 for communality we have the formula in the form—

$$h_i^2 = \frac{H_i^2 - q_i^2}{p_i^2} \quad (i = 2, 3, 4 \ldots)$$

The exception is the new communality of the trait or quality which has been *directly* selected, in our Example No. 1 the Stanford-Binet scores. For the directly selected trait the new communality is given by—

$$h_1^2 = \frac{p_1^2 H_1^2}{1 - q_1^2 H_1^2}$$

(Thomson, 1938*b*, 455; and see also Ledermann, 1938*b*). With these formulæ we can see what is likely to happen to a whole factorial analysis when the persons who are the subjects of the tests are only a sample of the wider population in which the analysis was first made.

4. *Hierarchical numerical example.*—We shall take, in the first place, the perfectly hierarchical example of our Chapters I and II. But to save space in the tables we shall consider only the first four tests. Their matrix of correlations, with the one common factor and the four specifics added, and with communalities inserted in the diagonal cells, was as follows:

	1	2	3	4	g	s_1	s_2	s_3	s_4
1	(·81)	·72	·63	·54	·90	·44	.	.	.
2	·72	(·64)	·56	·48	·80	.	·60	.	.
3	·63	·56	(·49)	·42	·70	.	.	·71	.
4	·54	·48	·42	(·36)	·60	.	.	.	·80
g	·90	·80	·70	·60	1·00	.	.	.	.
s_1	·44	.	.	.	.	1·00	.	.	.
s_2	.	·60	.	.	.	.	1·00	.	.
s_3	.	.	·71	.	.	.	.	1·00	.
s_4	.	.	.	·80	.	.	.	.	1·00

The bottom right-hand quadrant shows, by its zero entries, that the factors are all uncorrelated with one another, that is, orthogonal. The tests expressed as linear functions of the factors are—

$$z_1 = \cdot 9g + \cdot 436 s_1$$
$$z_2 = \cdot 8g + \cdot 600 s_2$$
$$z_3 = \cdot 7g + \cdot 714 s_3$$
$$z_4 = \cdot 6g + \cdot 800 s_4$$

These equations are only another way of expressing the same facts as are shown in the north-east, or the south-west, quadrant of the matrix (where only two places of decimals are used for the specific loadings, to keep the printing regular).

Let us now suppose that this matrix and these equations refer to a wide and defined population, e.g. all Massachusetts eleven-year-olds, and let us ask what will be the most likely matrix of correlations between these tests and factors to be found in a sample chosen by their scores in Test 1 so as to be more homogeneous. The variance of Test 1 in the wider population being taken as unity, let us take that in the more homogeneous select sample as being $p_1^2 = \cdot 36$. We then have, using $q_i = q_1 R_{1i}$, and treating g and the specifics just like tests, the following table:

	1	2	3	4	g	s_1	s_2	s_3	s_4
q	·80	·576	·504	·432	·720	·349	.	.	.
p	·60	·817	·864	·902	·694	·937	1	1	1
p^2 (variance)	·36	·668	·746	·813	·482	·878	1	1	1

For the correlations and communalities, using our formula—

$$\frac{R_{ij} - q_i q_j}{p_i p_j}$$

we get (again printing only two decimal places):

	1	2	3	4	g	s_1	s_2	s_3	s_4
1	(·61)	·53	·44	·36	·78	·28	.	.	.
2	·53	(·46)	·38	·31	·68	−·26	·73	.	.
3	·44	·38	(·32)	·26	·56	−·22	.	·83	.
4	·36	·31	·26	(·21)	·46	−·18	.	.	·89
g	·78	·68	·56	·46	1·00	−·39	.	.	.
s_1	·28	−·26	−·22	−·18	−·39	1·00	.	.	.
s_2	.	·73	.	.	.	.	1·00	.	.
s_3	.	.	·83	.	.	.	.	1·00	.
s_4	.	.	.	·89	.	.	.	.	1·00

In the more homogeneous sample, therefore, the correlations and the communalities of all the tests have sunk. The g column shows what the new correlations of g are with the tests; and on examination of the matrix we see that these, when cross-multiplied with one another, still give the rest of the matrix. Thus—

$$\cdot 78 \times \cdot 46 = \cdot 36 \ (r_{14})$$
$$\cdot 68^2 = \cdot 46 \ (h_2{}^2)$$

The test matrix is still of rank 1 (Thomson, 1938*b*, 458), and these g-column entries can become the diminished loadings of the single common factor required by Rank 1.

The columns for the specifics s_2, s_3 (and later specifics also) still show only one entry. In the bottom right-hand quadrant, zero entries show that these specifics are still uncorrelated with one another and with g, that is, g, s_2, s_3, and s_4 are still orthogonal.

But something has happened to the specific s_1. It has become correlated with g, and with all the tests. It has become an oblique factor, orthogonal still to the other specifics, but inclined to g and the tests. It leans further away from Test 1 than it formerly did, and makes obtuse angles (negative correlation) with the other tests and with g, to which it was originally orthogonal.

But since, as we have already pointed out, the test matrix with the reduced communalities is still of rank 1, it is clear that a fresh analysis could be made of the tests into one common factor and specifics, thus—

$$z_1' = \cdot 778g' + \cdot 628s_1'$$
$$z_2' = \cdot 679g' + \cdot 734s_2$$
$$z_3' = \cdot 562g' + \cdot 827s_3$$
$$z_4' = \cdot 462g' + \cdot 887s_4$$

In these equations the factors g', s_1', s_2, s_3, and s_4 are again orthogonal (uncorrelated), and the loadings shown give the correlations and give unit variances. This is the analysis which an experimenter would make who began with the sample and knew nothing about any test measurements in the whole population.

The reader, comparing the loadings in these equations

with the correlations in the matrix of the sample, will rightly conclude that the specifics from s_2 onward have not changed. In the matrix it is clear that they are still orthogonal, and their correlations with the tests, in the matrix, are the same as their loadings in the equations. The tests are, in the sample, more heavily loaded with these specifics than they were in the population, but the specifics are the same in themselves.

The new specific s_1' the reader will readily agree to be different from s_1. The latter became oblique in the sample, whereas s_1' is orthogonal. What now is to be said about the common factors g (in the population) and g' (in the sample)? From the fact that the loadings of g', in the sample equations, are identical with the correlations of the original g with the tests, in the sample matrix, one is tempted to imagine g' and g to be identical in nature. But that is not so certain.

If we go back to the equations of the tests in the population, we can rewrite them in the following form—

$$\begin{aligned}
z_1 &= \cdot467g' + \cdot800g'' + \cdot377s_1' \\
z_2 &= \cdot555g' + \cdot576g'' + \cdot600s_2 \\
z_3 &= \cdot485g' + \cdot504g'' + \cdot714s_3 \\
z_4 &= \cdot417g' + \cdot432g'' + \cdot800s_4
\end{aligned}$$

with two common factors g' and g'' instead of one common factor g. These equations still give the same correlations. For example—

$$r_{14} = \cdot467 \times \cdot417 + \cdot800 \times \cdot432 = \cdot540 \text{ as before.}$$

In these equations the specifics s_2, s_3, s_4 are the same, and the communalities of Tests 2, 3, and 4 are the same. All that we have done in these three tests is to divide the common factor g into two components. The ratio of the loading of g'' to the loading of g' is the same in each of them. The loadings of g'' we have made identical with the shrinkages q in the table on page 177.

In Test 1 also we have made the loading of g'' equal to the shrinkage $q_1 = \cdot8$. But in this test g'' cannot be looked upon merely as a component of g. To give the correct correlations, the loading of g' has to be $\cdot467$ as shown, and

the communality of Test 1 has been raised from its former value (·81) to—

$$\cdot 467^2 + \cdot 800^2 = \cdot 858$$

while the loading of the specific has correspondingly sunk. The factors g', g'', and s_1' are a totally new analysis of Test 1 in the population. Part of the former specific has been incorporated in the common factors.

Now let the factor g'' be abolished, i.e. held constant, so that the tests (now of less than unit variance, so we write them with x instead of z) are—

		Variances
$x_1 = \cdot 467g'$	$+ \cdot 377s_1'$	·360
$x_2 = \cdot 555g'$	$+ \cdot 600s_2$	·668
$x_3 = \cdot 485g'$	$+ \cdot 714s_3$	·746
$x_4 = \cdot 417g'$	$+ \cdot 800s_4$	·813

The reduced variances are the sum of the squares of the surviving loadings, e.g.—

$$\cdot 467^2 + \cdot 377^2 = \cdot 360$$

The variances, it will be seen, are the p^2's of our tests as measured in the sample. If each of the last set of equations is divided through by the square root of its variance, we arrive at the equations—

$$\begin{aligned} z_1' &= \cdot 778g' + \cdot 628s_1' \\ z_2' &= \cdot 679g' + \cdot 734s_2 \\ z_3' &= \cdot 562g' + \cdot 827s_3 \\ z_4' &= \cdot 462g' + \cdot 887s_4 \end{aligned}$$

which is the analysis already given as that of an experimenter who knew only the sample. As to the nature of g', we can say in Tests 2, 3, and 4 that it is possible to regard it as a component of the g of the population. But we cannot do so with assurance in Test 1. There its nature is more dubious. At all events, it is not the same common factor as in the population, and at best we can say that it is one of its components.

5. *A sample all alike in Test 1.*—These phenomena are still more striking if we consider a case where the sample is composed of persons who are all *alike* in Test 1. It would be an excellent exercise for the reader to calculate

the resulting matrix of correlations for tests and population factors in this case. The tests act in this case as though their original equations in the population had been—

$$\begin{aligned}
z_1 &= \qquad\qquad\quad g'' \\
z_2 &= \cdot 849g' + \cdot 720g'' + \cdot 600s_2 \\
z_3 &= \cdot 805g' + \cdot 630g'' + \cdot 714s_3 \\
z_4 &= \cdot 262g' + \cdot 540g'' + \cdot 800s_4
\end{aligned}$$

and then g'' had become zero, i.e. a constant with no variance.

It perhaps helps to a further understanding of what is happening to the factors during selection if we realize that holding the score of Test 1 constant does not hold its factors g and s_1 constant. They can vary in the sample from man to man, but since—

$$z_1 = \cdot 9g + \cdot 436s_1$$

remains constant, a man in the sample who has a high g must have a low s_1—that is, these factors are negatively correlated in the sample. And because they are thus negatively correlated, those members of the sample who have high g's, and who will therefore tend to do well in Tests 2, 3, and 4, will tend to have values below average (negative values) for their s_1, which will be therefore negatively correlated with these tests, in this sample.

So far in our examples we have assumed the sample to be more homogeneous than the population. But a sample can be selected to be less homogeneous. In such a case the same formulæ will serve, if we simply make the capital letters refer to the sample and the small to the population. In fact, the same tables, with their rôles reversed, can illustrate this case. In practical life we usually know which of two groups we would call the sample, and which the population. But mathematically there is no distinction, the one is a distortion of the other, and which is the " true " state of affairs is a question without meaning.

It must also throughout be remembered that all these formulæ and statements refer, not to consequences which are certain to follow, but to consequences which are to be expected. If actual samples were made the values experi-

mentally found in them for correlations, communalities, loadings, etc., would oscillate about those given by our formulæ, violently in the case of small samples, only slightly in the case of large samples.

6. *An example of rank* 2.—The above example has only one common factor. We turn next to consider an example with two. Again it is, we suppose, the first test, according to which the sample is deliberately selected, and again we suppose the "shrinkage" q_1 to be ·8. The matrices of correlations and communalities, in the population and in the sample, are then as follows, the two factors f_1 and f_2 and the specifics being treated in the calculation exactly as if they were tests. To economize room on the page, we omit the later specifics:

Correlations in the Population

	1	2	3	4	5	f_1	f_2	s_1	s_2
1	(·65)	·46	·59	·36	·41	·70	·40	·59	.
2	·46	(·37)	·36	·26	·23	·60	·10	.	·79
3	·59	·36	(·61)	·32	·45	·50	·60	.	.
4	·36	·26	·32	(·20)	·22	·40	·20	.	.
5	·41	·23	·45	·22	(·34)	·30	·50	.	.
f_1	·70	·60	·50	·40	·30	(1·00)	.	.	.
f_2	·40	·10	·60	·20	·50	.	(1·00)	.	.
s_1	·59	.	.	.	.	.	.	(1·00)	.
s_2	.	·79	.	.	.	.	.	.	(1·00)

Correlations in the Sample

	1	2	3	4	5	f_1	f_2	s_1	s_2
1	(·40)	·30	·40	·23	·26	·51	·25	·40	.
2	·30	(·27)	·23	·17	·12	·51	−·02	−·21	·85
3	·40	·23	·50	·22	·35	·32	·54	−·29	.
4	·23	·17	·22	(·13)	·14	·30	·12	−·16	.
5	·26	·12	·35	·14	(·26)	·15	·44	−·19	.
f_1	·51	·51	·32	·30	·15	(1·00)	−·23	−·36	.
f_2	·25	−·02	·54	·12	·44	−·23	(1·00)	−·18	.
s_1	·40	−·21	−·29	−·16	−·19	−·36	−·18	(1·00)	.
s_2	.	·85	.	.	.	.	.	.	(1·00)

We see here a new phenomenon. The two common factors f_1 and f_2 in the population were orthogonal to one another, as is shown by the zero correlation between them.

But in the sample they are negatively correlated (—·228); that is, they are oblique. We begin to see a generalization which can be algebraically proved, that *all the factors, common and specific, which are concerned with the directly selected test(s) become oblique to each other and to all the tests, but the specifics of the indirectly selected tests remain orthogonal* to everything, except each to its own test.

But *the matrix of the tests themselves is still of rank* 2, and an experimenter working only with the sample would find this out, although he would know nothing about the population matrix. He would therefore set to work to analyse it into two common factors, orthogonal to one another. A Thurstone analysis comes out in two common factors exactly, and can be rotated until all the loadings are positive. For example:

Test	1	2	3	4	5
Factor f_1'	·570	·521	·436	·332	·238
Factor f_2'	·276	.	·555	·130	·452

These factors f', however, are clearly a different pair from the factors f in the original population. In the sample, those original factors (f) are oblique; these (f') are orthogonal.

Again the whole phenomenon is reversible. The second matrix (with the orthogonal factors f') might refer to the population, and a sample picked with a suitable *increased* scatter of Variate 1. All our formulæ could be worked backwards, and we should arrive at the matrix beginning (·65), referring now to the sample. The f' factors would have become oblique, and a new analysis, suitably rotated, would give us the other factors f.

It becomes evident that the factors we obtain by the analysis of tests depend upon the subpopulation we have tested. They are not realities in any physical sense of the word; they vary and change as we pass from one body of men to another. It is possible, and this is a hope hinted at in Thurstone's book *The Vectors of Mind*, that if we could somehow identify a set of factors throughout all their changes from sample to sample (in most of which

they would be oblique) as being in some way unique, we might arrive at factors having some measure of reality and fixity. How Thurstone hopes to achieve this will be described in a later chapter. The work outlined in the present chapter, however, makes the writer far from optimistic that this can be achieved, or is even theoretically possible.

7. *A simple geometrical picture of selection.*—The geometrical picture of correlation between tests and factors which was described in Chapter IV is of some help in seeing exactly what happens to factors under selection in some test or trait. In Figure 23, x_1 represents the vector of the test or trait which is to be directly selected for, and g and s_1 are the axes of the common factor and of its specific—taking the case of one common factor only. The circle indicates the circular nature of the crowd of points which represent the population. It is a line of equal density of that crowd, which is densest at the origin and thins off equally in all directions. One quarter of that crowd are above average in both g and s_1, and another quarter are below average in both. The correlation between g and s_1 is zero.

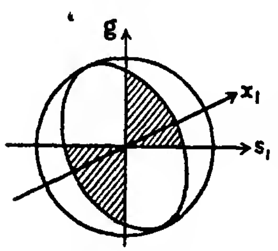

Figure 23.

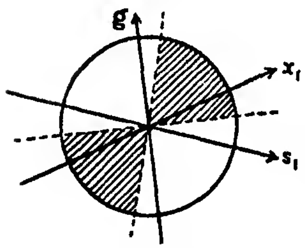

Figure 24.

But in the selected sample (Figure 23) the scatter of the persons along the test vector x_1 has been reduced. Persons have been removed from the whole crowd to leave the sample, but they have not been removed equally over the whole crowd. The line of equal density has become an ellipse, which is shorter along the line of the test vector x_1 than at right angles to that line. If we now compare the figure with Figure 18 in Chapter V (page 67), we see that it represents a state of negative correlation between g and s_1.

Less than one-quarter are now above average in both g and s_1, less than one-quarter below average in both. A majority are cases of being above average in the one factor and below in the other.

An experimenter coming first to the sample and knowing nothing about the population will naturally standardize each of his tests. He can do, indeed, nothing else. That is to say, he treats the crowd as again symmetrical and our ellipse as a circle (Figure 24). In *his* space, therefore, the lines g and s_1 will be at an obtuse angle, just as the axes in Section 2 of Chapter V became acute. He knows nothing about these lines, but chooses new axes for himself which are at right angles. One of these may be one of the old axes, but they cannot both coincide with the old axes.

8. *Random selection.*—These considerations, in Sections 1–7, deal with the results to be expected when a sample is deliberately selected so that the variance of one test is changed to some desired extent. The new variances and the changed correlations of the other tests given by our formula—

$$r_{ij} = \frac{R_{ij} - q_i q_j}{p_i p_j}$$

are not the *certain* result of our action in selecting for Test 1. If we selected a large number of samples of the same size, all with the same reduced variance in Test 1, they would not all be alike in the resulting correlations. On the contrary, they would all be different. But most of them would be *like* the expected set, few would depart widely from that; and the departures would be in both directions, some samples lying on the one side, others on the other side, of our expectation.

If now, instead of selecting samples which are all alike in the variance of one nominated test, we take a large number of *random* samples of the same size, what would we find? Among them would be a number which were alike in the variance of Test 1, and these in the other part of the correlation matrix would have values which varied round about those given by our formula. We could also pick out, instead of a set all alike in the variance of Test 1,

a different set all alike in the variance of Test 4, say; and these would have values in the remainder of the matrix oscillating about our formula, in which Test 4 would replace Test 1. In short, a complex family of random samples would show a *structure* among themselves such that if we fix any one variance the average of that array of samples obeys our formula.* Random sampling will not merely add an "error specific" to existing factors, it will make complex changes in the common factors.

* On the author's suggestion, Dr. W. Ledermann has since proved this conjecture analytically in a paper as yet unpublished. His results cover also the case of multivariate selection (see next chapter).

CHAPTER XII

THE INFLUENCE OF MULTIVARIATE SELECTION *

1. *Altering two variances and the covariance.*—In the preceding chapter we have discussed the changes which occur in the variances and correlations of a set of tests, and in their factors, when the sample of persons tested is chosen according to their performance in one of the tests: we are next going to see the results of picking our sample by their performances in more than one of the tests, first of all in two of them. Take again, the perfectly hierarchical example of the last chapter and of Chapters I and II. We must this time go as far as six tests in order to see all the consequences. The matrix of correlations of these tests and their factors will be simply an extension of that printed on page 176.

Now let us imagine a sample picked so that the variance of Test 1 and also that of Test 2 is intentionally altered, and further, their covariance (and hence their correlation) changed to some predetermined value.

It is at once clear that in these two directly selected tests the factorial composition will in general be changed —can indeed be changed to anything which is not incompatible with common sense and the laws of logic. What, however, will be the resulting sympathetic changes in the variances and covariances of the other tests of the battery ?

In Chapter XI we altered the variance of Test 1 from unity to ·36. The *consequent* diminution in variance to be expected in Test 2 was, as is shown on page 177, from unity to ·668, and the *consequent* change in correlation from ·72 to ·53. Here, however, let us pick our sample so that the variance of the second test is also diminished to ·36, and so that the correlation between them, instead of falling, ***rises*** to ·833. We have, that is to say, chosen

* **Thomson, 1937 ; Thomson and Ledermann, 1938.**

people for our sample who tend to be rather more alike than usual in these two test scores, as well as being closely grouped in each, an unusual but not an inconceivable sample. Natural selection (which includes selection by the other sex in mating) has no doubt often preferred individuals in whom two organs tended to go together, as long legs with long arms, and the same sort of thing might occur in mental traits. In terms of variance and covariance we have changed the matrix:

$$\begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 1{\cdot}00 & {\cdot}72 \\ 2 & {\cdot}72 & 1{\cdot}00 \end{array} = R_{pp}$$

to the matrix:

$$\begin{array}{c|cc} & 1 & 2 \\ \hline 1 & {\cdot}36 & {\cdot}30 \\ 2 & {\cdot}30 & {\cdot}36 \end{array} = V_{pp}$$

for $\dfrac{{\cdot}30}{\sqrt{({\cdot}36 \times {\cdot}36)}} = \dfrac{5}{6} = {\cdot}833$, the new correlation. Notice that the diagonal entries here (unities in R_{pp} and ·36, ·36 in V_{pp}) are the variances, not the communalities.

2. *Aitken's multivariate selection formula.*—We shall symbolically represent the whole original matrix of variances and covariances by:

$$\left[\begin{array}{c|c} R_{pp} & R_{pq} \\ \hline R_{qp} & R_{qq} \end{array}\right]$$

where the subscript p refers to the directly selected or picked tests, and the subscript q to all the other tests and the factors. R_{pq} (and also R_{qp}) means the matrix of covariances of the picked tests with all the others, including the factors. R_{qq} means the matrix of variances and covariances of the latter among themselves. Since at the outset the tests and factors are all assumed to be stan-

dardized, the variances in this whole R matrix are all unity, and the covariances are simply coefficients of correlation. In our case the R matrix is:

Analysis in the Population

	1	2	3	4	5	6	g	s_1	s_2	s_3	s_4	s_5	s_6
1	1·00	·72	·63	·54	·45	·36	·90	·44	.	.	.	.	.
2	·72	1·00	·56	·48	·40	·32	·80	.	·60	.	.	.	.
3	·63	·56	1·00	·42	·35	·28	·70	.	.	·71	.	.	.
4	·54	·48	·42	1·00	·30	·24	·60	.	.	.	·80	.	.
5	·45	·40	·35	·30	1·00	·20	·50	.	.	.	.	·87	.
6	·36	·32	·28	·24	·20	1·00	·40	.	.	.	.	.	·92
g	·90	·80	·70	·60	·50	·40	1·00	.	.	.	.	.	.
s_1	·44	.	.	.	.	.	.	1·00	.	.	.	.	.
s_2	.	·60	.	.	.	.	.	.	1·00	.	.	.	.
s_3	.	.	·71	.	.	.	.	.	.	1·00	.	.	.
s_4	.	.	.	·80	.	.	.	.	.	.	1·00	.	.
s_5	.	.	.	.	·87	.	.	.	.	.	.	1·00	.
s_6	.	.	.	.	.	·92	.	.	.	.	.	.	1·00

The R_{pp} matrix is the square 2×2 matrix, the R_{qq} matrix the square 11×11 matrix, while R_{pq} has two rows and eleven columns, R_{qp} being the same transposed.

Our object is to find what may be expected to happen to the rest of the matrix when R_{pp} is changed to V_{pp}. Formulæ for this purpose were first found by Karl Pearson, and were put into the matrix form in which we are about to quote them by A. C. Aitken (Aitken, 1934). The matrix changes to:

$$\begin{array}{c|c} V_{pp} & V_{pp} R_{pp}^{-1} R_{pq} \\ \hline R_{qp} R_{pp}^{-1} V_{pp} & R_{qq} - R_{qp} (R_{pp}^{-1} - R_{pp}^{-1} V_{pp} R_{pp}^{-1}) R_{pq} \end{array}$$

and in order to explain the meaning of these formulæ we shall carry out the calculation for a part of the above matrix only (the first four tests), with a strong recommendation to the reader to perform the whole calculation systematically. If we confine ourselves to the first four tests we have—

$$R_{pp} = \begin{bmatrix} 1·00 & ·72 \\ ·72 & 1·00 \end{bmatrix}$$

$$R_{qq} = \begin{bmatrix} 1·00 & ·42 \\ ·42 & 1·00 \end{bmatrix}$$

$$R_{pq} = \begin{bmatrix} {\cdot}68 & {\cdot}54 \\ {\cdot}56 & {\cdot}48 \end{bmatrix}$$

$$R_{qp} = \begin{bmatrix} {\cdot}63 & {\cdot}56 \\ {\cdot}54 & {\cdot}48 \end{bmatrix}$$

3. *The calculation of a reciprocal matrix.*—The most tiresome part of the calculation, if the number of directly selected tests is large, is to find R_{pp}^{-1} the reciprocal of the matrix R_{pp}. By the reciprocal of a matrix is meant another matrix such that the product—

$$R_{pp} \,.\, R_{pp}^{-1} = \begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix} = I$$

where I is the so-called "unit matrix" which has unit entries in the diagonal and zero entries everywhere else. Such a reciprocal matrix can be found by means of Aitken's method of pivotal condensation as follows (Aitken, 1937*a*). Write the given matrix with the unit matrix below it and minus the unit matrix on its right, thus:

				Check Column
1·00	·72	−1·0000	.	·72
·72	1·00	.	−1·0000	·72
1·00	.	.	.	1·00
.	1·00	.	.	1·00
	·4816	·7200	−1·0000	·2016
	1·0000	1·4950	−2·0764	·4186
	− ·7200	1·0000	.	·2800
	1·0000	.	.	1·0000
		2·0764	−1·4950	·5814
		−1·4950	2·0764	·5814
		R_{pp}^{-1}		

As before, we divide the first row of each slab through by its first member, writing the result in a row left blank for that purpose. Each pivot is thus unity, the whole calculation is made easier, and the process continues until the left-hand column no longer has any contents, when the numbers in the middle column are the reciprocal matrix. For large matrices the advantages of this automatic form of calculation are more pronounced. That the matrix is

indeed the reciprocal we can check by direct calculation. We have—

$$R_{pp} \,.\, R_{pp}^{-1} = \begin{bmatrix} 1{\cdot}00 & {\cdot}72 \\ {\cdot}72 & 1{\cdot}00 \end{bmatrix} \begin{bmatrix} 2{\cdot}0764 & -1{\cdot}4950 \\ -1{\cdot}4950 & 2{\cdot}0764 \end{bmatrix} = \begin{bmatrix} 1 & . \\ . & 1 \end{bmatrix}$$

Matrix multiplication is carried out by obtaining the inner products (see footnote, page 81) of the *rows* of the first matrix with the *columns* of the second. Thus—

$$1 \times 2{\cdot}0764 - {\cdot}72 \times 1{\cdot}4950 = 1$$
$$-1 \times 1{\cdot}4950 + {\cdot}72 \times 2{\cdot}0764 = 0$$

are the two upper entries in the product matrix. When the reciprocal matrix R_{pp}^{-1} has thus been calculated, the best way of proceeding is to find—

$$C = R_{pp}^{-1}\, R_{pq}$$

and
$$D = R_{qq} - R_{qp}\, C$$

In the case of our example these are—

$$C = \begin{bmatrix} 2{\cdot}0764 & -1{\cdot}4950 \\ -1{\cdot}4950 & 2{\cdot}0764 \end{bmatrix} \begin{bmatrix} {\cdot}63 & {\cdot}54 \\ {\cdot}56 & {\cdot}48 \end{bmatrix} = \begin{bmatrix} {\cdot}4709 & {\cdot}4037 \\ {\cdot}2209 & {\cdot}1894 \end{bmatrix}$$

$$D = \begin{bmatrix} 1{\cdot}00 & {\cdot}42 \\ {\cdot}42 & 1{\cdot}00 \end{bmatrix} - \begin{bmatrix} {\cdot}63 & {\cdot}56 \\ {\cdot}54 & {\cdot}48 \end{bmatrix} \begin{bmatrix} {\cdot}4709 & {\cdot}4037 \\ {\cdot}2209 & {\cdot}1894 \end{bmatrix}$$

$$= \begin{bmatrix} 1{\cdot}00 & {\cdot}42 \\ {\cdot}42 & 1{\cdot}00 \end{bmatrix} \quad \begin{bmatrix} {\cdot}4204 & {\cdot}3604 \\ {\cdot}3604 & {\cdot}3089 \end{bmatrix}$$

$$= \begin{bmatrix} {\cdot}5796 & {\cdot}0596 \\ {\cdot}0596 & {\cdot}6911 \end{bmatrix}$$

subtraction of matrices being carried out by subtracting each element from the corresponding one. We next need—

$$V_{pp}C = \begin{bmatrix} {\cdot}36 & {\cdot}30 \\ {\cdot}30 & {\cdot}36 \end{bmatrix} \begin{bmatrix} {\cdot}4709 & {\cdot}4037 \\ {\cdot}2209 & {\cdot}1894 \end{bmatrix} = \begin{bmatrix} {\cdot}2358 & {\cdot}2022 \\ {\cdot}2208 & {\cdot}1893 \end{bmatrix}$$

which gives us the new covariances of the directly selected tests with those indirectly selected. For V_{qq} we need still $C'(VC)$ where the prime indicates that the matrix is transposed (rows becoming columns)—

$$C'(VC) = \begin{bmatrix} {\cdot}4709 & {\cdot}2209 \\ {\cdot}4037 & {\cdot}1894 \end{bmatrix} \begin{bmatrix} {\cdot}2358 & {\cdot}2022 \\ {\cdot}2208 & {\cdot}1893 \end{bmatrix} = \begin{bmatrix} {\cdot}1598 & {\cdot}1370 \\ {\cdot}1370 & {\cdot}1175 \end{bmatrix}$$

and then—

$$V_{qq} = D + C'VC = \begin{bmatrix} {\cdot}5796 & {\cdot}0596 \\ {\cdot}0596 & {\cdot}6911 \end{bmatrix} + \begin{bmatrix} {\cdot}1598 & {\cdot}1370 \\ {\cdot}1370 & {\cdot}1175 \end{bmatrix}$$

$$= \begin{bmatrix} {\cdot}7394 & {\cdot}1966 \\ {\cdot}1966 & {\cdot}8086 \end{bmatrix}$$

We now can write down the whole new 4 × 4 matrix of variances and covariances. In the same way, had we included the other tests and the factors, we would have arrived at the whole new 13 × 13 matrix for all the variances and covariances which we now print.* The values calculated above for the first four tests will be recognized in its top left-hand corner. (The diagonal entries are variances, not communalities.)

Covariances in the Sample

	1	2	3	4	5	6	g	s_1	s_2	s_3	s_4	s_5	s_6
1	·36	·30	·24	·20	·17	·14	·34	·13	·05	.	.	.	.
2	·30	·36	·22	·19	·16	·13	·32	·04	·18	.	.	.	.
3	·24	·22	·74	·20	·16	·13	·33	−·14	−·07	·71	.	.	.
4	·20	·19	·20	·81	·14	·11	·28	−·12	−·06	.	·80	.	.
5	·17	·16	·16	·14	·87	·09	·23	−·10	−·05	.	.	·87	.
6	·14	·13	·13	·11	·09	·92	·19	−·08	−·04	.	.	.	·92
g	·34	·32	·33	·28	·23	·19	·47	−·19	−·10	.	.	.	.
s_1	·13	·04	−·14	−·12	−·10	−·08	−·19	·70	·32	.	.	.	.
s_2	·05	·18	−·07	−·06	−·05	−·04	−·10	·32	·43		.	.	.
s_3	.	.	·71	.	.	.	.	.	.	1·00	.	.	.
s_4	.	.	.	·80	.	.	.	.	.	.	1·00	.	.
s_5	.	.	.	.	·87	.	.	.	.	.	.	1·00	.
s_6	.	.	.	.	.	·92	.	.	.	.	.	.	1·00

4. *Features of the sample covariances.*—Examination of this matrix shows the following features :

(1) The specifics of the indirectly selected tests have remained unchanged. They are still orthogonal to each other and all the other tests and factors (except each to its own test), are still of unit variance, and have still the same covariances with their own tests, though these will become larger *correlations* when the tests are restandardized ;

(2) The specifics of the directly selected tests have become oblique common factors, correlated with everything except the other specifics ;

* In such calculations on a larger scale, the methods of Aitken's (1937*a*) paper are extremely economical. Triple products of matrices of the form $XY^{-1}Z$ can thus be obtained in one pivotal operation (see Appendix, paragraph 12 ; and Chapter VIII, Section 7, page 189).

(3) The matrix of the indirectly selected tests is still of the same rank (here rank 1);

(4) The variances of the factors g, s_1, and s_2 have been reduced to ·47, ·70, and ·43.

An experimenter beginning with this sample, and knowing nothing about the factors in the wider population, would have no means of knowing these relative variances, and would no doubt standardize all his tests. He certainly would not think of using factors with other than unit variance. And even if he were by a miracle to arrive at an analysis corresponding to the last table, with three oblique general factors, he would reject it (*a*) because of the negative correlations of some of the factors, and (*b*) because he can reach an analysis with only two common factors, and those orthogonal. It is therefore practically certain that he will not reach the population factors, at least as far as the directly selected tests are concerned. His data and his analysis will be as follows. The variances are all made unity and the covariances converted into correlations. The analysis into factors is a new one, not derived from the last table.

Analysis in the Sample

	1	2	3	4	5	6	g'	h	s_1'	s_2'	s_3	s_4	s_5	s_6
1	1·00	·83	·46	·38	·30	·24	·82	·45	·35	.	.	.	.	.
2	·83	1·00	·43	·35	·28	·22	·77	·45	.	·46	.	.	.	.
3	·46	·43	1·00	·26	·21	·16	·56	.	.	.	·83	.	.	.
4	·38	·35	·26	1·00	·17	·13	·46	.	.	.	.	·89	.	.
5	·30	·28	·21	·17	1·00	·11	·37	.	.	.	.	.	·93	.
6	·24	·22	·16	·13	·11	1·00	·29	.	.	.	.	.	.	·96
g'	·82	·77	·56	·46	·37	·29	1·00	.	.	.	.	.	.	.
h	·45	·45	.	.	.	.	.	1·00	.	.	.	.	.	.
s_1'	·35	.	.	.	.	.	.	.	1·00	.	.	.	.	.
s_2'	.	·46	.	.	.	.	.	.	.	1·00	.	.	.	.
s_3	.	.	·83	.	.	.	.	.	.	.	1·00	.	.	.
s_4	.	.	.	·89	.	.	.	.	.	.	.	1·00	.	.
s_5	.	.	.	.	·93	.	.	.	.	.	.	.	1·00	.
s_6	.	.	.	.	.	·96	.	.	.	.	.	.	.	1·00

5. *Appearance of a new factor.*—The most noticeable change in this sample analysis, as compared with the

population analysis on page 189, is the appearance of a new "factor" h linking the directly selected tests, a factor which is clearly due entirely to that selection. What degree of reality ought to be attributed to it? Does it differ from the other factors really, or have they also been produced by selection, even in the population, which is only in its turn a sample chosen by natural selection from past generations?

Otherwise the analysis is still into one common factor and specifics. The loadings of the common factor are less than they were in the population, and this, as our table of variances and covariances shows, is due to a real diminution in the variance of the common factor. The new common factor g' is a component of the old one.

The loadings of s_1 and s_2 have also sunk, because they have been in part turned into a new common factor. The loadings of the other specifics have risen. But this is entirely because the variance of the tests has sunk due to the shrinkage in g, and is not due to any new specifics being added.

All these considerations make it very doubtful indeed whether any factors, and any loadings of factors, have absolute meaning. They appear to be entirely dependent upon the population in which they are measured, and for their definition there would be required not only a given set of tests and a given technical procedure in analysis, but also a given population of persons.

In our example, the covariance of Tests 1 and 2 in the new matrix V_{pp} was made larger than would naturally follow from the changed variances of Tests 1 and 2, so that the correlation increased. In consequence the new factor h is one with positive loadings in both tests.

We might equally well, however, have decreased the covariance in V_{pp}, for example making—

$$V_{pp} = \begin{bmatrix} \cdot 86 & \cdot 04 \\ \cdot 04 & \cdot 86 \end{bmatrix}$$

and in that case (the reader is strongly recommended to carry out the calculations as an exercise) the new factor h will be an interference factor, with negative loading in one

of the two tests. In this case the experimenter, with a dislike for such negative loadings, would probably " rotate " his factors away from any position which had any simple relation to the factors of the population.

Again, the formulæ, moreover, can all be worked backward, the sample treated as the population and the population as the sample; though as we said before, samples in real life are certainly, as a rule, more homogeneous in nearly every quality than the complete population.

PART IV

CORRELATIONS BETWEEN PERSONS

CHAPTER XIII

REVERSING THE RÔLES *

1. *Exchanging the rôles of persons and tests.*—In all the previous chapters the correlations considered have been correlations between tests, and the experiments envisaged were experiments in which comparatively few tests were administered to a large number of persons. For each test there would, therefore, be a long list of marks. The whole set of marks would make an oblong matrix, with a few *rows* for the tests, and a very large number of *columns* for the persons—we will choose that way of writing it, of the two possibilities.

From such a set of marks we then calculated the correlation coefficients for each pair of tests, and our analysis of the tests into factors was based upon these. In the process of calculating a correlation coefficient we do such things to the row of marks in each test as finding its average, and finding its standard deviation. We quite naturally assume that we can legitimately carry out these operations. We assume, that is, that in the row of marks for one test these marks are comparable magnitudes which at any rate rise and fall with some mental quality even if they do not strictly speaking measure it in units, like feet or ounces.

The question we are going to ask in this part of this book is whether, in the above procedure, the rôles of persons and of tests can be exchanged (Thomson, 1935*b*, 75, Equation 17), and if so what light this throws upon

* The first explicit references to correlations between persons in connexion with factor technique seem to have been made independently and almost simultaneously by Thomson (1935*b*, July) and Stephenson (1935*a*, August), the former being pessimistic, the latter optimistic. But such correlations had actually been used much earlier by Burt and by Thomson, and almost certainly by others, probably without full consciousness of their special interest.

factorial analysis. Instead of comparatively few tests (perhaps two or three dozen; fifty-seven is the largest battery reported up to date) and a very large number of persons, suppose we have comparatively few persons, and a large number of tests, and find the correlations between the persons. In that case our matrix of marks would be oblong in the other direction, with a large number of rows for the tests, and a small number of columns for the persons, and each correlation, instead of being as before between two rows, would be between two columns. Taking only small numbers for purposes of an explanatory table, we would have in the ordinary kind of correlations a table of marks like this:

	Persons						
	×	×	×	×	×	×	×
Tests	×	×	×	×	×	×	×
	×	×	×	×	×	×	×

while for correlations between persons we would have a table of marks like this:

	Persons		
	×	×	×
	×	×	×
	×	×	×
Tests	×	×	×
	×	×	×
	×	×	×
	×	×	×

But we meet at once with a serious difficulty as soon as we attempt to calculate a correlation coefficient between two persons from the second kind of matrix. To do so, we must find the average of each column, just as previously we found the average of each row for the other kind of correlation. But to find the average of each column (by adding all the marks in that column together and dividing by their number) is to assume that these marks are in some sense commensurable up and down the column, although each entry is a mark for a different test, on a scoring system which is wholly arbitrary in each test (Thomson, 1935*b*, 75–6).

To make this difficulty more obvious, let us suppose that the first four tests are:

1. A form-board test;
2. A dotting test;
3. An absurdities test;
4. An analogies test.

In each of these the experimenter has devised some kind of scoring system. Perhaps in the form-board test he gives a maximum of 20 points, and in the dotting test the score may be the number of dots made in half a minute. But to find the average of such different things as this is palpably absurd, and the whole operation can be entirely altered by an arbitrary change like taking the number of seconds to solve the form board instead of giving points.

2. *Ranking pictures, essays, or moods.*—This is a very fundamental difficulty which will probably make correlations between persons in the general case impossible to calculate. In certain situations, however, it does not arise, namely where each person can put the "tests" in an order of preference according to some criterion or judgment (Stephenson, 1935*b*), and it is with cases of this kind that we shall deal in the first place. Usually the "tests" here are not really different tests like those named above, but are perhaps a number of children's essays which have to be placed in order of merit, or a number of pictures in order of æsthetic preference, or a number of moods which the subject has to number, indicating the frequency of their occurrence in himself. Indeed, the subject might not only give an order of preference to, say, the essays, but might give them actual marks, and there would be no absurdity in averaging the column of such marks, or in correlating two such columns, made by different persons.

Such a correlation coefficient would show the degree of resemblance between the two lists of marks given to the children, or given to a set of pictures according to their æsthetic value. It would indicate, therefore, a resemblance between the minds of the two persons who marked the essays or judged the pictures. A matrix of correlations between several such persons might look exactly like the matrices of correlations between tests which occur in

Parts I and II, and could be analysed in any of the same ways. What would the "factors" which resulted from such an analysis mean when the correlations were between persons? Take an imaginary hierarchical case first.

3. *The two sets of equations.*—In test analysis the common factor found was taken to be something called into play by each test, the different tests being differently loaded with it. The test was represented by an equation such as—

$$z_4 = \cdot 6g + \cdot 8s_4$$

For each of the numerous persons who formed the subjects of the testing, an estimate was made of *his* g, and another estimate could be made of *his* s_4. The different tests were combined into a weighted battery for this purpose of estimating a man's amount of g. His score in Test 4 would then be made up of his g and s_4 inserted in the above specification equation.

$$z_{4 \cdot 9} = \cdot 6g_9 + \cdot 8s_{4 \cdot 9}$$

would be the score of the ninth person in Test 4.

By analogy, when we analyse a matrix consisting of correlations between persons, we arrive at a set of equations describing the persons in terms of common and specific factors. Corresponding to a hierarchical battery of tests, we could conceivably have a hierarchical team of persons, from which we would exclude any person too similar to one already included. Each person in the hierarchical team would then be made up of a factor he shared with everyone else in the team, and a specific factor which was his own idiosyncrasy. An equation like—

$$z_9 = \cdot 4g' + \cdot 917s_9'$$

would now specify the composition of the ninth person. g' is something all the persons have, s_9' is peculiar to Person 9. The *loadings* now describe the person, and the amount of g' "possessed" or demanded by each test can be estimated by exactly the same techniques employed in Part I. The score which Test 4 would elicit from Person 9 would be obtained by inserting the g' and s_9' "possessed"

by that test into the specification equation of Person 9, giving—

$$z_{9\cdot 4} = \cdot 4g_4' + \cdot 917s_{9\cdot 4}'$$

This equation is to be compared with the former equation—

$$z_{4\cdot 9} = \cdot 6g_9 + \cdot 8s_{4\cdot 9}.$$

Both equations ultimately describe the same score, but $z_{9\cdot 4}$ is not identical with $z_{4\cdot 9}$. The raw score X is the same, but the one standardized z is measured from a different zero, and in different units, from the other. Disregarding this for the moment, we see that with the exchange of rôles of tests and persons, *the loadings and the factors have also changed rôles.* Formerly, persons possessed different amounts of g, and tests were differently loaded with it. Now, tests possess different amounts of g', and persons are differently loaded with it. We feel impelled to inquire further into the relationships of these complementary factors and loadings.

The test which is most highly saturated with g is that one which, in terms of Spearman's imagery, requires most expenditure of general mental energy, and is least dependent upon specific neural engines. It correlates more with its fellow-members of the hierarchical battery than any other test among them does. It represents best what is common to them all.

The man, in a hierarchical team of men, who is most highly saturated with g' is that man who is most like all the others. His correlations with them are higher than is the case for any other man in the team. He is the individual who best represents the type. But a nearer approach to the type can be made by a weighted team of men, just as formerly we weighted a battery of tests to estimate their common factor.

4. *Weighting examiners like a Spearman battery.*—Correlations of this kind between persons were used long before any idea of what Stephenson has called " inverted factorial analysis " was present. The author and a colleague found in the winter of 1925–6 a number of correlations between experienced teachers who marked the essays written by

fifty schoolboys upon "Ships" (Thomson and Bailes, 1926). One table or matrix of such correlations, between the class teacher and six experienced head masters who marked the essays independently of one another, was as follows:

	Te	*A*	*B*	*C*	*D*	*E*	*F*
Te	.	·60	·69	·56	·69	·63	·67
A	·60	.	·53	·50	·54	·55	·68
B	·69	·53	.	·60	·65	·66	·64
C	·56	·50	·60	.	·67	·67	·65
D	·69	·54	·65	·67	.	·54	·69
E	·63	·55	·66	·67	·54	.	·69
F	·67	·68	·64	·65	·69	·69	.

In the article in question, these different markers were compared by correlating each with the pool of all the rest. These correlations are shown in the first row of the table below.

Purely as an illustrative example, let us make also an approximate analysis of this matrix, and take out at any rate its chief common factor. On the assumption that it is roughly hierarchical, we can use Spearman's formula—

$$\text{Saturation} = \sqrt{\left\{\frac{A^2 - A'}{T - 2A}\right\}}^{*}$$

More easily, we can insert its largest correlation coefficient as an approximate communality for each test, and find Thurstone's approximate first-factor loadings (see Chapter II, page 24). We get for the saturations or loadings the second and third rows of this table:

	Te	*A*	*B*	*C*	*D*	*E*	*F*
Correlation with pool of rest	·77	·67	·76	·73	·76	·75	·82
Spearman saturations	·814	·704	·796	·766	·798	·788	·861
Thurstone method	·81	·73	·80	·78	·80	·80	·85

We see that *F* is the most "typical" examiner of these essays, in the sense that he is more highly saturated with what is common to all of them; while *A* conforms least to the herd.

With the same formula which in Part I we used to esti-

* See Chapter IX, page 154.

mate a man's g from his test-scores, we could here estimate an essay's g' from its examiner scores. That is to say, the marks given by the different examiners would be weighted in proportion to the quantities—

$$\frac{\text{Saturation with } g'}{1 - \text{saturation}^2}$$

where g' is that quality of an essay which makes a common appeal to all these examiners. Their marks (after being standardized) would therefore be weighted in the proportions $\cdot 814/(1 - \cdot 814^2)$, etc., that is:

	Te	*A*	*B*	*C*	*D*	*E*	*F*
	2·41	1·40	2·17	1·85	2·20	2·08	3·33
or	·72	·42	·65	·56	·66	·63	1·00

to make global marks for the essays, which could then be reduced to any convenient scale. If this were done, the result would be the " best " estimate * of that aspect or set of aspects of the essay which all these examiners are taking into account, disregarding all that can possibly be regarded as idiosyncrasies of individual examiners. Whether we think it the best estimate in other senses is a matter of subjective opinion. We may wish the " idiosyncrasies " (the specific, that is) of a certain examiner to be given great weight. It clearly would not do, for example, to exclude Examiner *A* from the above team *merely* because he is the most different from the common opinion of the team, without some further knowledge of the men and the purpose of the examination. The " different " member in a team might, for example, be the only artist on a committee judging pictures, or the only Democrat in a court judging legal issues, or the only woman on a jury trying an accused girl. But in non-controversial matters, if all are of about equal experience, it is probable that this system of weighting, restricting itself to what is certainly common to all, will be most generally acceptable as fairest.

* Best whether we adopt the regression principle or Bartlett's. For if only one " common factor " is estimated, the difference is one of unit only, and the weighting in the text is the " best " on both systems.

5. *Example from "The Marks of Examiners."*—This form of weighting examiners' marks has probably never yet been used in practice. But it has been employed, by Cyril Burt, in an inquiry into the marks given by examiners (Burt, 1936). As an example, we take the marks given independently by six examiners to the answer papers of fifteen candidates aged about 16, in an examination in Latin. (The example is somewhat unusual, inasmuch as these candidates were a specially selected lot who had all been adjudged equal by a previous examiner, but it will serve as an illustration if the reader will disregard that fact.) The marks were (*op. cit.*, 20):

Cand.	*A*	*B*	*C*	*D*	*E*	*F*	*Examiners*
1	39	43	52	37	43	40	
2	39	44	50	43	43	46	
3	44	51	55	47	46	46	
4	37	46	43	44	40	43	
5	38	47	55	35	43	45	
6	45	50	54	45	45	49	
7	42	52	51	45	44	46	
8	48	49	53	47	46	46	
9	32	42	49	34	36	38	
10	37	40	48	37	39	42	
11	38	42	47	39	36	39	
12	40	44	50	41	36	42	
13	38	43	50	36	34	41	
14	35	45	49	37	40	40	
15	32	38	41	28	34	34	

The correlations between the examiners calculated from this table are (the examiner with the highest total correlation leading):

	F	*A*	*B*	*E*	*D*	*C*
F	.	·86	·84	·82	·84	·71
A	·86	.	·80	·74	·85	·71
B	·84	·80	.	·80	·81	·67
E	·82	·74	·80	.	·72	·69
D	·84	·85	·81	·72	.	·48
C	·71	·71	·67	·69	·48	.

If, assuming this table to be hierarchical, we find each examiner's saturation with the common factor by Spear-

man's formula, we obtain (with Professor Burt, *op. cit.*, 294):

F	A	B	E	D	C
·95	·92	·91	·87	·84	·72

In the sense, therefore, of being most typical, F is here the best examiner. The proportionate weights to be given to each examiner, in making up that global mark for the candidate which will best agree with the common factor of the team of examiners, are, as before—

$$\frac{\text{Saturation}}{1 - \text{saturation}^2}$$

provided the marks have first been standardized. The resulting weights, giving F the weight unity, are:

F	A	B	E	D	C
1·00	·61	·54	·37	·29	·15

(If the weights are to be applied to the raw or unstandardized marks, they must each be divided by that examiner's standard deviation.)

The marks thus obtained are only an estimate of the "true" common-factor mark for each child, just as was the case in estimating Spearman's g; and the correlation of these estimates with the "true" (but otherwise undiscoverable) mark will be, as there (Chapter VII, page 106)—

$$r_m = \sqrt{\frac{S}{1 + S}}$$

where S is the sum of all the six quantities—

$$\frac{\text{Saturation}^2}{1 - \text{saturation}^2}$$

In our case this gives—

$$r_m = ·98$$

The best examiner's marking itself correlated with the hypothetical "true" mark to the amount ·95, so that the improvement is not worth the trouble of weighting, especially as the simple average of the team of examiners gives ·97. But in some circumstances the additional labour might be worth while, and there is an interest in

knowing which examiners conform least and which most to the team, and having a measure of this.

After the saturation of each examiner with the hypothetical common factor has been found, the correlations due to that factor can be removed from the table exactly as in analysing tests in Chapter II, pages 27 and 28, or in Chapter IX, page 155. The residues, as there, may show the presence of other factors; and "specific" resemblances or antagonisms between pairs of examiners, or minor factors running through groups of examiners, may be detected and estimated.

In short, all the methods of Parts I and II of this book there used on correlations between tests may be employed on correlations between examiners. The tests have come alive and are called examiners, that is all. But since the child's performance, judged by the different examiners differently, is here nevertheless the same identical performance, our interpretation of the results is different. The two cases throw light on one another. A Spearman hierarchical battery of tests may estimate each child's general intelligence, which is there something in common among the tests. The examiners may have been instructed to mark exclusively for what they think is general intelligence. In that case their weighted team will estimate for each child a general intelligence, which is something in common among the somewhat discrepant ideas the examiners hold on this matter.

6. *Preferences for school subjects.*—In the previous sections we have discussed correlations between examiners who all mark the same examination papers. The purpose of their marking these papers is to award prizes, distinctions, passes, and failures to the candidates. The examiners are a means to this end; the reason for employing several of them is to obtain a list of successes and failures in which we can have greater confidence. The technique described is one which enables us to combine their marks, on certain assumptions, to greatest advantage. But it can, as in the inquiries described in *The Marks of Examiners*, be turned to compare individual examiners, and to evaluate the whole process of examining.

It is only a step to another, very similar, experiment in which objects evaluated by the "examiners" are not the works of candidates in an examination, but are objects chosen for the express purpose of gaining an insight into the minds of those asked to judge them. Thus we might ask several persons each to evaluate on some scale the æsthetic appeal of forty or fifty works of art (Stephenson, 1936*b*, 353), or ask a number of school pupils each to place in order of interest a list of school subjects.

Stephenson (1936*a*) asked forty boys and forty girls attending a higher school in Surrey, England, thus to place in order of their preference twelve school subjects represented by sixty examination papers, and calculated for about half these pupils the correlation coefficients between them. To explain the kind of outcome that may be expected from such an experiment it will be sufficient for us to quote his data for a smaller number of pupils, say eight girls, avoiding anomalous cases for simplicity in a first consideration. The correlations between them were as follows (*op. cit.*, 50):

Girl	3	4	5	7	17	18	19	20
3	.	·59	·31	·26	−·02	−·16	−·38	−·35
4	·59	.	·75	·42	−·23	−·01	−·66	−·03
5	·31	·75	.	·65	−·29	−·02	−·18	−·08
7	·26	·42	·65	.	−·50	−·15	−·54	−·17
17	·02	−·23	−·29	−·50	.	·60	·52	·72
18	−·16	−·01	−·02	−·15	·60	.	·09	·79
19	−·38	−·66	−·18	−·54	·52	·09	.	·40
20	−·35	−·03	−·08	−·17	·72	·79	·40	.

This table at once suggests that these girls fall into two types. Girls 3, 4, 5, and 7 correlate positively among themselves; they have somewhat similar preferences among school subjects. Girls 17, 18, 19, and 20 correlate positively among *them*selves. But the two groups correlate negatively with one another. The two types were different in their order of preference, Type I tending, for example, to put English and French higher, and Physics and Chemistry lower, than Type II (though both were agreed that Latin was about the least lovable of their studies !).

7. *A parallel with a previous experiment.*—This experiment, it will be seen, forms a parallel to that inquiry (also by Stephenson) described in Chapter I, Section 9, where *tests* fell into two types, verbal and pictorial, with correlations falling there as here into four quadrants. If we call the two types of school pupil here the linguistic (L) and the scientific (S), and again use C for the cross-correlations, the diagram corresponding to that on page 16 of Chapter I is:

L	C
C	S

The chief difference between the two cases is that there the cross-correlations, though smaller than hierarchical order in the whole table would demand, were nevertheless positive. Here, however, the cross-correlations are actually negative.

It is true that the signs of all the correlations in the C quadrants can in either case be reversed, by reversing the order of the lists either of all the earlier or all the later variables (there tests, here pupils). But that is not really permissible in either case. We have no doubt which is the top and which the bottom end of a list of marks, whether in a verbal test or a pictorial test; and to reverse the order of preference given by either the linguistic or the scientific pupils would be simply to stultify the inquiry. There is, therefore, a real difference between the cases. In the present set of correlations something is acting as an "interference factor."

In Chapter I we explained the correlations and their tetrad-differences by the hypothesis of three uncorrelated factors g, v, and p required in various proportions by the tests, and possessed in various amounts by the children. The *loadings* which indicated the proportions of the factors in each test we tacitly assumed to be all positive. Thur-

stone expressly says that it is contrary to psychological expectation to have more than occasional negative loadings.

8. *Negative loadings.*—Let us endeavour to make at least a qualitative scheme of factors to express the correlations between the pupils, factors possessed in various amounts by the subjects of the school curriculum, and demanded in various proportions by each pupil before he will call the subject interesting. One type of pupil weights heavily the linguistic factor in a subject in evaluating its interest to him. The other type weights heavily the scientific factor in a subject in judging its attraction for him. But to explain actual negative correlations between pupils we must assume that some of the loadings are negative, assume, that is, that some of the children are actively repelled by factors which attract others. Common sense does not think thus. Common sense says that two children may put the subjects in opposite orders, even though they both like them all, provided they don't like them equally well. But then common sense is not anxious to analyse the children into *uncorrelated additive* factors. If each child is thus expressed as the weighted sum of various factors, two children can correlate negatively only if some of the loadings are negative in the one child and positive in the other, for the correlation is the inner product of the loadings. Since Stephenson has found numerous negative correlations between persons, and since few negative correlations are reported between tests, we seem here to have an experimental difference between the two kinds of correlation, and if ever correlations between persons come to be analysed as minutely and painstakingly as correlations between tests, it would seem that the free admission of negative loadings would be necessary.* The present matrix can in fact be roughly analysed into two general factors, one of which has positive loadings in all pupils, while the other is positively loaded in the one type, negatively loaded in the other.

9. *An analysis of moods.*—A still more ingenious application by Stephenson of correlations between persons is in an experiment in which for each person a "population"

* See Stephenson, 1936*b*, 349.

of thirty *moods*, such as "irascible," "cheerful," "sunny," were rated for their prevalence and intensity for each of ten patients in a mental hospital, and for six normal persons (Stephenson, 1936*c*, 363). This time the correlation table indicated three types, corresponding to the manic-depressives, the schizophrenes, and the normal persons, each type correlating positively within itself, but negatively or very little with the other types. These experiments were only illustrative, and it remains to be seen whether factors which will prove acceptable psychologically will be isolated in persons in the same manner as *g*, and the verbal factor, have been isolated in tests. The parallel between the two kinds of correlation and analysis is, however, certainly likely to throw light on the nature of factors of both kinds.

CHAPTER XIV

THE RELATION BETWEEN TEST FACTORS AND PERSON FACTORS

1. *Burt's example, centred both by rows and by columns.*—In the examples we have just considered, there is no doubt that correlations between persons can be calculated without absurdity. In the matrix of marks given by a number of examiners (marking the same paper) to a number of candidates, either two candidates can be correlated, or two examiners. The heterogeneity of marks referred to in Chapter XIII, Section 1, does not enter as a difficulty. Still keeping to such material, let us ask ourselves what the relation is between factors found in the one way, and factors found in the other. Qualitatively, we have already suggested that factors and loadings change rôles in some manner. The most determined attempt to find an exact relationship has been that made by Cyril Burt, who concludes that, if the initial units have been suitably chosen, the factors of the one kind of analysis are identical with the loadings of the other, and vice versa (Burt, 1937*b*). The present writer, while agreeing that this is so in the very special circumstances assumed by Burt, is of opinion that his is a very narrow case, and that the factors considered by Burt are not typical of those in actual use in experimental psychology. Theoretically, however, Burt's paper is of very great interest. It can be presented to the general reader best by using Burt's own small numerical example, based on a matrix of marks for four persons in three tests:

	Persons	*a*	*b*	*c*	*d*
	1	-6	2	0	4
Tests	2	3	1	-1	-3
	3	3	-3	1	-1

It will be noticed that this matrix of marks is already centred both ways. The rows add up to zero, and so do the columns. The test scores have been measured from their means, and then thereafter the columns of personal scores have been measured from their means; or it can be done persons first, tests second, the end result being the same. Burt does not give the matrix of raw scores from which the above matrix comes.

If we take the doubly centred matrix as he gives it, the matrices of variances and covariances formed from it are:

Test Covariances

	1	2	3
1	56	− 28	− 28
2	− 28	20	8
3	− 28	8	20

Person Covariances

	a	*b*	*c*	*d*
a	54	− 18	0	− 36
b	− 18	14	− 4	8
c	0	− 4	2	2
d	− 36	8	2	26

Notice that in both these matrices the columns add to zero, just as they do in the matrices of residues in the "centroid" process.

2. *Analysis of the covariances.*—Burt next proceeds to analyse each of these by Hotelling's method. It seems clear that there will exist some relation between the two analyses, since the primary origin of each matrix is the same table of raw marks, and to show that relation most clearly Burt *analyses the covariances direct*, and not the correlations which could be made from each table (by dividing each covariance by the square root of the product of the two variances concerned). For the two Hotelling analyses he obtains (and the Thurstone factors before rotation would here be the same):

Analysis of the Tests

$$x_1 = 2\sqrt{14}\,\gamma_1$$
$$x_2 = -\sqrt{14}\,\gamma_1 + \sqrt{6}\,\gamma_2$$
$$x_3 = -\sqrt{14}\,\gamma_1 - \sqrt{6}\,\gamma_2$$

Analysis of the Persons

$$\begin{aligned} a &= -3\sqrt{6}f_1 \\ b &= \sqrt{6}f_1 + 2\sqrt{2}f_2 \\ c &= \quad\quad - \sqrt{2}f_2 \\ d &= 2\sqrt{6}f_1 - \sqrt{2}f_2 \end{aligned}$$

In both cases *two* factors are sufficient (there will always be fewer Hotelling or Thurstone factors than tests with a doubly centred matrix of marks, for a mathematical reason). The reader can check that the inner products give the covariances, e.g.—

covariance $(bd) = \sqrt{6} \times 2\sqrt{6} - 2\sqrt{2} \times \sqrt{2} = 12 - 4 = 8$

The method of finding Hotelling loadings was described in Chapter V, and the reader can readily check that the coefficients of γ_1, for example, do act as required by that method. For if we use numbers proportional to $2\sqrt{14}$, $-\sqrt{14}$, and $-\sqrt{14}$, namely 1, $-\frac{1}{2}$, $-\frac{1}{2}$, as Hotelling multipliers we get :

56	− 28	− 28	1
− 28	20	8	− ½
− 28	8	20	− ½

56	− 28	− 28
14	− 10	− 4
14	− 4	− 10
84	− 42	− 42

proportional to 1 − ½ − ½ as required.

The largest total (84) is the first "latent root," and the multipliers 1, $-\frac{1}{2}$, $-\frac{1}{2}$, have to be divided, according to Chapter V, by the square root of the sum of their squares, and multiplied by the square root of 84, giving—

$$2\sqrt{14} \quad -\sqrt{14} \quad -\sqrt{14}$$

3. *Factors possessed by each person and by each test.*—Burt then goes on to "estimate," by "regression equations," the amount of the factors γ possessed by the persons, and the amount of the factors f possessed by the tests. There is a misuse of terms here, for with Hotelling factors there is no need to "estimate"; they can be accurately calculated: but that is a small point. The first three equations can be solved for the γ's—there is indeed one equation too many, but it is consistent. And the four equations of the second group can be solved for the f's—again they are consistent. Since the equations are consistent, we can choose the easiest pair in each case to solve for the two unknowns. Choosing the two equations for x_1 and x_2 we obtain—

$$\gamma_1 = \frac{1}{2\sqrt{14}} x_1$$

$$\gamma_2 = \frac{x_2 + \frac{1}{2}x_1}{\sqrt{6}}$$

For the other set of factors we naturally choose the equations in a and c, and have—

$$f_1 = -\frac{a}{3\sqrt{6}}$$

$$f_2 = -\frac{c}{\sqrt{2}}$$

Now, since we are very liable to confusion in this discussion, let us remind ourselves what these factors γ and these factors f are. The factors γ are factors into which each test has been analysed. They do not vary in amount from test to test, but each test is differently loaded with them. They vary in amount from person to person.

The factors f are factors into which each person has been analysed. These do not vary in amount from person to person, but from test to test. Each person is differently loaded with them, that is, made up of them in different proportions. The γ's are uncorrelated fictitious tests: the f's are uncorrelated fictitious persons.

Now, from the equations—

$$\gamma_1 = \frac{1}{2\sqrt{14}} x_1$$

$$\gamma_2 = \frac{x_2 + \frac{1}{2}x_1}{\sqrt{6}}$$

we can find the amount of each factor γ_1 and γ_2 possessed by each person, by inserting his scores x_1 and x_2 in these equations, scores which are given in the matrix:

	a	*b*	*c*	*d*
1	− 6	2	0	4
2	3	1	− 1	− 3
3	3	− 3	1	− 1

Thus the first person possesses γ_1 in an amount $-6/2\sqrt{14}$, because his x_1 is − 6. For the four persons and the two factors we find the amounts of these factors possessed by each person to be:

Factors	γ_1	γ_2
a	$-\frac{3}{\sqrt{14}}$	0
b	$\frac{1}{\sqrt{14}}$	$\frac{2}{\sqrt{6}}$
c	0	$-\frac{1}{\sqrt{6}}$
d	$\frac{2}{\sqrt{14}}$	$-\frac{1}{\sqrt{6}}$

4. *Reciprocity of loadings and factors.*—These are the ***amounts*** of the factors γ possessed by the four persons. If now the reader will compare them with the ***loadings*** of the factors f in the second set of equations on page 215, he will see a resemblance. The signs are the same, and the zeros are in the same places. Moreover, the resemblance becomes identity if we destandardize the factors f_1 and f_2, measuring the former in units $\sqrt{84}$ times as large, and the latter in units $\sqrt{12}$ times as large, 84 and 12 being the

non-zero latent roots of both matrices. In these units let us use ϕ_1 and ϕ_2 for them. The equations on page 215 giving the analysis of the persons then become—

$$
\begin{aligned}
a &= \frac{-3\sqrt{6}}{\sqrt{84}}(\sqrt{84}f_1) &&= -\frac{3}{\sqrt{14}}\phi_1\\
b &= \frac{\sqrt{6}}{\sqrt{84}}(\sqrt{84}f_1) + \frac{2\sqrt{2}}{\sqrt{12}}(\sqrt{12}f_2) &&= \frac{1}{\sqrt{14}}\phi_1 + \frac{2}{\sqrt{6}}\phi_2\\
c &= \qquad\qquad\qquad - \frac{\sqrt{2}}{\sqrt{12}}(\sqrt{12}f_2) &&= \qquad\quad - \frac{1}{\sqrt{6}}\phi_2\\
d &= \frac{2\sqrt{6}}{\sqrt{84}}(\sqrt{84}f_1) - \frac{\sqrt{2}}{\sqrt{12}}(\sqrt{12}f_2) &&= \frac{2}{\sqrt{14}}\phi_1 - \frac{1}{\sqrt{6}}\phi_2
\end{aligned}
$$

It will be seen that the loadings of ϕ_1 and ϕ_2 are identical with the amounts of γ_1 and γ_2 in the table on page 217. A similar calculation could be made comparing the amounts of f_1 and f_2 possessed by the tests with the loadings of γ_1 and γ_2 (suitably destandardized) in the analysis of the tests. As we said at the outset, if suitable units are chosen for the marks and the factors, the *loadings* of the personal equations are the *factors* of the test equations, and the *factors* of the personal equations are the *loadings* of the test equations. *But only for doubly centred matrices of marks.* It would be wrong to conclude in general that loadings and factors are reciprocal in persons and tests.

Indeed, even for doubly centred matrices of marks, this simple reciprocity holds only for the analysis of the covariances and not for analyses of the matrices of correlations. Except by pure accident (and as it happens, Burt's example is in the case of test correlations such an accident), the saturations of the correlation analysis will not be any *simple* function of the loadings of the covariance analysis.

5. *Special features of a doubly centred matrix.*—But in any case, a matrix of marks which has been centred both ways is one in which only a very special kind of residual association between the variables is present. Most of what we commonly call the association or resemblance between either tests or persons, the amount of which we gauge by the correlation coefficient, is due to something over and

above this. We can write down an infinity of possible raw matrices from which Burt's doubly centred matrix might have come. To the rows of the latter matrix we can add *any quantities we like* without in the slightest altering the correlations between the tests, but making enormous changes in the correlations between the persons. Let us, for example, add 10 to the top row, 18 to the middle row, and 16 to the bottom row. There results the matrix:

	a	b	c	d	
1	4	12	10	14	
2	16	14	12	10	(A)
3	19	18	17	15	

This gives as correlations between the persons:

	a	b	c	d
a	1·00	·75	·84	—·14
b	·75	1·00	·28	—·76
c	·84	·28	1·00	·42
d	—·14	—·76	·42	1·00

Next, without changing this matrix of correlations between persons in the slightest, we can add *any quantities we like* to the columns of the matrix of marks, and produce an infinity of different matrices of correlations between tests. If, for example, we add 5, 2, 8, and 9 to the four columns, we have a matrix of raw marks:

	a	b	c	d	
1	9	14	18	23	
2	21	16	20	19	(B)
3	24	15	25	24	

This has the same correlations between persons, but the correlations between tests are now:

	1	2	3
1	1·00	—·16	·24
2	—·16	1·00	·92
3	·24	·92	1·00

Or instead, by adding suitable numbers to the columns and to the rows, we might have arrived at the matrix:

	a	*b*	*c*	*d*	
1	44	48	18	10	
2	63	57	27	13	(*C*)
3	58	48	24	10	

or equally well at:

	a	*b*	*c*	*d*	
1	35	45	37	43	
2	34	34	26	26	(*D*)
3	34	30	28	28	

The order of merit of the persons in each test is quite different in each of these matrices. The order of difficulty of the tests for each person is quite different in each. If we consider the ordinary correlation between Tests 1 and 2, we find that it is negative in (*B*), zero in (*D*), and positive in (*C*), yet all of these matrices reduce to Burt's matrix when centred both ways. It is clear that they contain factors of correlation which are absent in the doubly centred matrix.

The averages of the rows and the columns of (*C*) are as follows:

	a	*b*	*c*	*d*	*Average*
1	44	48	18	10	30
2	63	57	27	13	40
3	58	48	24	10	35
Average	55	51	23	11	

The correlation between two tests is clearly influenced very much by the fact that here the person *a* is so much cleverer than the person *d*. Similarly, the correlation between two persons is influenced by the fact that Test 1 is more difficult than Test 2. As soon as the matrix is centred both ways, all the correlation due to these and similar influences is almost extinguished. Centred by rows, (*C*) becomes:

$$\begin{bmatrix} 14 & 18 & -12 & -20 \\ 23 & 17 & -13 & -27 \\ 23 & 13 & -11 & -25 \end{bmatrix}$$

and all the tests are equally difficult on the average. Centred by columns *as well*, it becomes :

$$\begin{matrix} -6 & 2 & 0 & 4 \\ 3 & 1 & -1 & -3 \\ 3 & -3 & 1 & -1 \end{matrix}$$

and not only are all the tests equally difficult on the average, but all the persons are equally clever on the average. It is to the covariances still remaining that Burt's theorem about the reciprocity of factors and loadings applies. It does not apply to the full covariances of the matrix centred only one way, in the manner usually meant when we speak of covariances or of correlations.

6. *Profile correlations.*—The correlations calculated from such doubly centred matrices might, the present writer suggests, be termed " profile correlations." They remind one of, but are not in general identical with, partial correlations for constant average scores in a certain set of tests (or persons). They depend in an intricate way on the other tests (or persons) in the battery or team, since the centring depends on what other scores are present. The name " *profile* correlations " is suggested because the correlation between two tests, say, is dependent upon the profile of the two rows of scores, after a general " handicapping " of all persons to the same average in the battery: and similarly the profile correlation between two persons is the resemblance between them in a battery of tests after all the tests have been " handicapped " to the same average level of difficulty over the persons.

In Figure 25 the left-hand portion illustrates the full correlation between Tests 1 and 2 in matrix (C) centred by rows only. The correlation coefficient of :

$$\frac{1,824}{\sqrt{(1,064 \times 1,716)}} = \cdot 98$$

is mainly due to the fact that both curves come steeply downhill from *a* to *d*, to the fact, that is, that the four persons differ considerably in average ability.

The right-hand portion of Figure 25 represents the *profile* correlation between these two tests, in this particular battery. The correlation due to *a* being clever and *d*

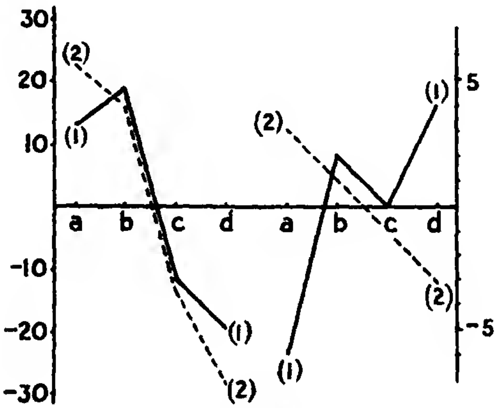

Figure 25

stupid, etc., has been removed. What remains is a negative correlation, and its negative sign reflects the fact that, in the left-hand diagram, Test 1 begins below and finishes above Test 2. A profile correlation might have been made by equalizing the four persons on these two tests alone. Instead, they have been equalized on the battery of four tests.

PART V

THE INTERPRETATION OF FACTORS

CHAPTER XV

THE DEFINITION OF *g*

1. *Any three tests define a* " *g*."—This concluding part will be devoted to an attempt to answer the questions: " What *are* factors? What is their psychological and physiological interpretation? On what principles are we to decide between the different possible analyses of tests (and persons)? " It may seem strange to have deferred these considerations so long, and to have discussed methods of analysing tests, and of estimating factors, before asking explicitly what they mean. But that is how " factors " have arisen. Whatever else they are, they certainly are not things which can be identified with clearness first, and discussed and measured afterwards. Their definition and interpretation arise out of the attempt to measure them. We shall begin by discussing, in the present chapter, the definition and nature of *g*.

It will be remembered that the idea of *g* arose out of Professor Spearman's acute observation that correlation coefficients between tests tend to show hierarchical order: that is that their tetrad-differences tend to be zero or small; or in more technical terms still, that the rank to which a matrix of correlation coefficients can be " reduced " by suitable diagonal elements tends towards rank *one*. This fundamental fact is at the basis of all those methods of factorial analysis which magnify specific factors, and a reason for it, based on the idea that it is a mathematical result of the laws of probability, will be advanced in Chapter XVIII. In consequence of this fundamental fact, *correlation coefficients* between a number of variables can be adequately accounted for by a few common factors. To be adequately described by one only—a *g*—the " reduced " rank of the correlation matrix has to be *one*, within the limits of sampling error.

This trouble of sampling error is very liable to obscure

the issue, and we will remove it during most of the present chapter, as we did in Parts I and II, by supposing that we have defined our population (say all adult Scots, or all men, for that matter) and have tested every one of them.

Suppose now that we have three tests and have, in this whole population, measured their correlation coefficients:

	1	2	3
1	1	r_{12}	r_{13}
2	r_{12}	1	r_{23}
3	r_{13}	r_{23}	1

If, as is usually the case, these coefficients are all positive, and if each of them is at least as large as the product of the other two, we can explain them by assuming one g and three specifics s_1, s_2, and s_3. There are many other ways of explaining them, but let us adopt this one. *We have thereby defined a factor* g *mathematically* (Thomson, 1935*a*, 260). It is then for the psychologist to say, from a consideration of the three tests which define it, what name this factor shall bear and what its psychological description is. The psychologist may think, after studying the tests, that they do not seem to him to have anything in common, or anything worth naming and treating as a factor. That is for him to say. Let us suppose that at any rate he does not reject the possibility, but that he would like an opportunity of studying other tests which (mathematically speaking) contain this factor, and have nothing else in common, before finally deciding.

In that case the experimenter must search for a fourth test which, when added to these three, gives tetrad-differences which are zero; and then for a fifth and further tests, each of which makes zero tetrad-differences with the tests of the pre-existing battery. This extended battery the experimenter would lay before the psychological judge, to obtain a ruling whether the single common factor, of which it is the now extended but otherwise unaltered definition, is worthy of being named as a psychological factor.

2. *The extended or purified hierarchical battery.*—Mathematically, *any* three tests with which the experimenter

cared to begin would define " a " *g*, if we except temporarily the case, to which we shall later return, of three correlation coefficients, one of which is less than the product of the other two. The experimental tester, however, might in some cases have great difficulty in finding further tests, to add to the original three, which would give zero tetrad-differences. Unless he could do so, it is unlikely that the psychological judge would accept the factor as worthy of a name and separate existence in his thoughts. It is, for example, an experimental fact that starting with three tests which a general consensus of psychological opinion would admit to have only " intelligence " as a common requirement, it has proved possible to extend the battery to comprise about a score of tests without giving any tetrad-differences which cannot be regarded as zero. Even that has not been accomplished without difficulty, and without certain blemishes in the hierarchy having to be removed by mathematical treatment. But the fact that with these reservations it is possible, and that psychological judgment endorses the opinion that each test of this battery requires " intelligence," is the main evidence behind the actual " existence " of such a factor as " *g*, general intelligence." It must be noted that the word " existence " here does not mean that any physical entity exists which can be identified with this *g*. It does mean, however, that, as far as the experimental evidence goes, there is some aspect of the causal background which acts " as if " it were a single unitary factor in these tests.

The process of making such a battery of tests to define general intelligence (see Brown and Stephenson, 1933) has not in fact taken the form of choosing three tests as the basal definition and then extending the battery. Instead, a number of tests which, it was thought from previous experience, would act in the desired way have been taken, and the battery thus formed has then been purified by the removal of any tests which broke the hierarchy. The removal of such tests does not, of course, mean that they do not contain *g*, but it means that *g* is not their only link with the other tests of the battery, and that therefore they are unsuitable members of a set of tests intended to define *g*.

Further, the actual making of such a hierarchical battery has not been accomplished under the ideal conditions which we have been assuming, namely, that the whole population has been accurately tested. There always remains some doubt, therefore, whether, without the blurring effect of sampling error, the hierarchy would continue to be near enough to perfection. But these details should not be allowed to obscure the simplicity of the main argument. The important point to note is that the experimenter has produced a battery of tests which is, he claims, hierarchical; that the mathematician assures him that such a battery acts "as if" it had only one factor in common (though it *can* also be explained in many other ways), and that the psychologist, who may be the same person as the experimenter, agrees that psychologically the existence of such a factor as the sole link in this battery seems a reasonable hypothesis.

3. *Different hierarchies with two tests in common.*—Now, it must be remembered that, starting with three other tests, which may contain two of the former set, it may very well be possible to build up a different hierarchy. Only experiment could show whether this were possible in each case, there is no mathematical difficulty in the way. Such a hierarchy would also define "a" g, but this would be usually a different factor from the former g. If there were *three* tests common to the two hierarchies, then the two g's could be identified with one another (sampling errors apart), and the three tests would be found to have the same saturations with the one g as with the other. But if only two tests were common to the two batteries this would not in general be the case, and the different saturations of these tests with the two g's would show that the latter were different (Thomson, 1935*a*, 261–2). Under such circumstances the psychologist has to choose. He cannot have both these g's. Both are mathematically of equal standing, it is a psychological decision which has to be made. When one g is accepted, the other, as a factor, must then be rejected and a more complicated factorial analysis of the second hierarchy has to be built up which is consistent with this. A simple artificial example will

illustrate this. Suppose that four tests give a perfect hierarchy of correlations thus:

	1	2	3	4
1	1·00	·72	·63	·54
2	·72	1·00	·56	·48
3	·63	·56	1·00	·42
4	·54	·48	·42	1·00

On the principle that the smallest possible number of common factors must be chosen, the analysis of these tests would be—

$$\left.\begin{aligned} z_1 &= \cdot 9g + \sqrt{\cdot 19}\,s_1 \\ z_2 &= \cdot 8g + \sqrt{\cdot 36}\,s_2 \\ z_3 &= \cdot 7g + \sqrt{\cdot 51}\,s_3 \\ z_4 &= \cdot 6g + \sqrt{\cdot 64}\,s_4 \end{aligned}\right\} \quad (A)$$

Suppose now that Tests 2 and 4 are brigaded with two other tests, 5 and 6, in a new experiment, and that the correlations found are:

	2	4	5	6
2	1·00	·48	·42	·54
4	·48	1·00	·56	·72
5	·42	·56	1·00	·63
6	·54	·72	·63	1·00

This is also a perfect hierarchy, and the principle of parsimony in common factors leads to the analysis—

$$\left.\begin{aligned} z_2 &= \cdot 6g' + \sqrt{\cdot 64}\,t_2 \\ z_4 &= \cdot 8g' + \sqrt{\cdot 36}\,t_4 \\ z_5 &= \cdot 7g' + \sqrt{\cdot 51}\,t_5 \\ z_6 &= \cdot 9g' + \sqrt{\cdot 19}\,t_6 \end{aligned}\right\} \quad (B)$$

But this analysis is inconsistent with the former, for the saturations of z_2 and z_4 with their common link have changed. If the factor g has been accepted as a psychological entity, then the factor g' cannot be. To be consistent we must begin our equations for z_2 and z_4 in the

same manner as before, and although we may split up their specifics to link them with the new tests, the only link between them themselves must be g. We can then complete the analysis in various ways,* of which one is—

$$\begin{aligned} z_3 &= \cdot 8g \quad + \cdot 6s_3 \\ z_4 &= \cdot 6g \quad + \cdot 529150h + \sqrt{\cdot 36}t_4 \\ z_5 &= \cdot 525g + \cdot 468006h + \sqrt{\cdot 51}t_5 \\ z_6 &= \cdot 675g + \cdot 595294h + \sqrt{\cdot 19}t_6 \end{aligned}$$

4. *A test measuring "pure* g."—Although the hierarchical battery defines a g, it does not enable it to be measured exactly (but only to be estimated) unless either it contains an infinite number of tests, or a test can be found which conforms to the hierarchy and has a g saturation of unity.† In the latter case this test which is "pure g" is such that when it is considered along with any other two tests of its hierarchy, its correlations with them, multiplied together, give the intercorrelation of those two with one another: if k is the "pure" test, then—

$$r_{ik}r_{jk} = r_{ij}$$

its g saturation being—

$$\sqrt{\frac{r_{ik}r_{jk}}{r_{ij}}} = 1$$

No such "pure" test of the g which is defined by the Brown-Stephenson hierarchy of nineteen tests has yet been found. Such a pure test, with full g saturation, must not be confused with tests which are sometimes called tests of pure g because they do not contain certain other factors, in particular the verbal factor. Thus the "S.V.P."

* Four tests are insufficient as a defining battery for two common factors.

† It is understood, of course, that even such a test would give different measures of a man's g from day to day, if the man's performance in it varied (as it undoubtedly would) from day to day. By measuring with exactness is meant, in this part of the text, measurement free from the uncertainty due to the factors outnumbering the tests. The reader is reminded that we are assuming sampling errors to be nil, the whole population having been tested.

(Spearman Visual Perception) tests are referred to by Dr. Alexander (1935, 48) as a " pure measure of g " ; but their saturations with g are given by him (page 107) as ·757, ·701, and ·736 respectively, so that in each case only about half the variance is " g." A possible alternative to the plan of first defining g and then seeking to improve its estimate would be to *begin* with three tests satisfying the relation—

$$r_{ik}r_{jk} = r_{ij}$$

which were reasonably acceptable as a definition of general intelligence, and give greater content to the psychological significance of *this* g by discovering tests which were hierarchical with these three. The lack of an exact measure of what is at present called g is a serious practical defect. Another possible way of remedying this will be referred to below in connexion with what are there called " singly conforming " tests. First, however, let us consider the case where three tests are such that—

$$r_{ik}r_{jk} > r_{ij}$$

5. *The Heywood case.*—In such a case the g saturation of the test k, if we calculate it, is greater than unity, which is impossible. Yet it is possible, in theory at least, to add tests to such a triplet to form an extended hierarchy with zero tetrad-differences. There can be one such case (but only one) in a hierarchy. We shall call them *Heywood cases*, as this possibility was first pointed out by him (Heywood, 1931). As an artificial example consider these correlations :

	1	2	3	4	5
1	1·000	·945	·840	·735	·630
2	·945	1·000	·720	·630	·540
3	·840	·720	1·000	·560	·480
4	·735	·630	·560	1·000	·420
5	·630	·540	·480	·420	1·000

This is a perfect hierarchy, every tetrad-difference being exactly zero. It is, moreover, a perfectly possible set of correlations, and passes the tests required for a matrix of correlations to be possible. For example, the determinant of the matrix is positive (see Chapter IV, Section 3, page

58). But when we calculate the g saturations of the tests we find them to be:

Test	1	2	3	4	5
g saturation	1·05	·9	·8	·7	·6

so that a single general factor is an impossible explanation of this hierarchy as far as Test 1 is concerned. The correlations of Test 1 with the other tests are possible, and they give exactly zero tetrad-differences: but yet the test cannot be a "two-factor" test, for the correlations of the first row are too high to be explained in that way.

We might well have possessed the hierarchy of Tests 2, 3, 4, and 5 first, before we discovered Test 1. We should then have analysed these four as follows in a two-factor analysis—

$$
\begin{aligned}
z_2 &= \cdot 9g + \cdot 436s_2 \\
z_3 &= \cdot 8g + \cdot 600s_3 \\
z_4 &= \cdot 7g + \cdot 714s_4 \\
z_5 &= \cdot 6g + \cdot 800s_5
\end{aligned}
$$

We then, let us suppose, discover Test 1, with its impossible g saturation. We want to retain the above analysis for the other tests. Now can we analyse Test 1 to explain its correlations with them? We can do so in several ways. If we give it arbitrarily the loading ·955 for g, we must use the specific of each test to give the additional correlation required. We thus arrive at the following possible but complicated analysis of Test 1—

$$z_1 = \cdot 955g + \cdot 196s_2 + \cdot 127s_3 + \cdot 098s_4 + \cdot 071s_5 + \cdot 141s_1$$

Here Test 1 is seen as containing each of the specifics of the four other tests, and only a small specific loading of its own. We have used up nearly all its variance in explaining the correlations. Clearly there must be a limit to this process. If another test were added to the hierarchy, we might entirely exhaust the available variance of Test 1 in explaining its correlations. Or, indeed, the reader might add, we might more than exhaust it, and prove the impossibility of adhering to the pre-existing analysis. But this is not so. Such a test would only prove *the impossibility of its own existence*, if we may make

an Irish bull. Suppose, for example, a Test 6 were to turn up with the correlations :

	1	2	3	4	5
6	·882	·756	·672	·588	·504

Such a test, when brigaded with Tests 2, 3, 4, and 5, would be given the analysis—

$$z_6 = \cdot 84g + \cdot 533s_6$$

If now we use even the whole specific of this test as a link with Test 1, we cannot explain the correlation ·882. We would need for that a loading of ·150 for s_6 in Test 1, and we have not enough variance left in Test 1 for this. But when this happens, we find that we have allowed the matrix of correlations to become an impossible one. If we add Test 6 to our matrix and calculate its determinant, we find it negative, which cannot occur in practice. The Test 6 could not occur, if the previous five tests already existed. Or vice versa, if Tests 2–6 existed, the Heywood case given would be impossible. The rule governing its possible existence has been given by Ledermann, namely, that the g saturation of the Heywood case cannot exceed—

$$\sqrt{\frac{1 + S}{S}}$$

where S is the quantity familiar from Spearman's formula—

$$S = \Sigma \frac{r_{ig}^2}{1 - r_{ig}^2}$$

for the *remainder* of the hierarchy ($i = 2, 3, 4 \ldots$). If, then, we have a large hierarchy, we shall find it impossible to discover a test which conforms to it and which at the same time has a g saturation greater than unity. If we have a small hierarchy containing a Heywood case, we shall find it impossible to discover many tests to add to it, except indeed by the formal device of adding tests which do not correlate with it at all. All these considerations make it appear likely that if a Heywood test can be found to conform to a hierarchy, the g defined by that hierarchy must be abandoned. The seeker for a test for pure g is thus in a delicate position. He wants to find a test with

full saturation of unity. But he must just hit the mark. If the saturation exceeds unity, his whole hierarchy must be abandoned as a definition. And even when the exact saturation of unity has been found, there seems to be too narrow a line dividing the perfect from the impossible, and the reality of the g seems to be balanced on a knife edge. In actual practice, of course, sampling errors would make the situation less acute and could for some time be called in to explain a certain amount of excess saturation over unity.

6. *Hierarchical order when tests equal persons in number.*— If a test cannot be found whose saturation with g is unity (" pure g "), the other method of measuring g exactly would seem to be to extend the hierarchy until it comprised so many tests that the multiple correlation with g—

$$r_m = \sqrt{\frac{S}{S+1}}$$

became practically unity. For S increases with the number of tests, being the sum of the positive quantities—

$$\frac{r_{ig}^2}{1 - r_{ig}^2}$$

There is here a point of some theoretical interest, namely, what happens when we have increased the number of hierarchical tests until they are as numerous as the persons to whom they are given ? This, in view of the difficulty of finding tests to add to a hierarchy, is admittedly not a question likely to trouble experimenters, but its theoretical implications are considerable.

It can be shown that whenever we have a matrix of correlations based upon the same number of tests as persons, its determinant is zero. Now the determinant of a hierarchical matrix (with unity in each diagonal cell) can be shown to be of the form—

$$\begin{aligned}
&(1 - r_{1g}^2)(1 - r_{2g}^2)(1 - r_{3g}^2)(1 - r_{4g}^2) \ldots \\
+ &\qquad\quad r_{1g}^2\,(1 - r_{2g}^2)(1 - r_{3g}^2)(1 - r_{4g}^2) \ldots \\
+ &(1 - r_{1g}^2)\qquad r_{2g}^2\,(1 - r_{3g}^2)(1 - r_{4g}^2) \ldots \\
+ &(1 - r_{1g}^2)(1 - r_{2g}^2)\qquad r_{3g}^2\,(1 - r_{4g}^2) \ldots \\
+ &(1 - r_{1g}^2)(1 - r_{2g}^2)(1 - r_{3g}^2)\qquad r_{4g}^2 \ldots \\
+ &\ldots
\end{aligned}$$

and it is clear that each of these quantities is positive unless we have a case of pure g, or a Heywood case. A case of pure g will leave one of the rows of the above sum non-zero. To make the whole sum zero, one case must be a Heywood case, giving—

$$1 - r_{ig}^2 \text{ negative.}$$

It would seem, therefore, that by the time we have added hierarchical tests to make them equal in number to the persons, we will necessarily have added a Heywood hierarchical case (of which there can be only one in a hierarchy). But we have agreed that the discovery of a Heywood case will cause us to abandon the hierarchy as a definition of g!

Mathematically this seems to mean that although the quantity S increases with each new test, provided it is not a Heywood case, yet S does not increase indefinitely, and the multiple correlation does not converge to perfect correlation.

The case discussed above, where the number of tests is increased to equal the number of persons, may seem to the reader to be an academic case only. But the case of reducing the number of persons until they equal the number of tests is one which could easily be realized in practice, and presents equal theoretical difficulties. This draws attention from a new point of view to what has already been emphasized in Part III, the dependence of any definition of factors on the sample of persons tested. If we have a perfect hierarchy of, say, 50 tests, in a population of, say, 1,000 persons, and we reduce the number of persons by discarding some at random, it is, of course, to be expected that the correlations will change, and the hierarchy become disturbed. It would, however, at first sight appear possible to discard them so skilfully as not to disturb the hierarchy, or at least not disturb it much. But it would seem from the above considerations that try as we might, we could not, as the number of persons decreased towards fifty, prevent the correlations changing so as to give us a Heywood case, if we clung to hierarchical order. Or to put the same point in another way: a

sample of fifty persons from the above thousand, if it gives hierarchical order, will give a Heywood case, and its *g* will be impossible.

If the *g* corresponding to the original analysis on the thousand persons were anything real, such as a given quantity of mental energy available in each person, then it ought always to be possible, one might erroneously think, to find fifty persons and fifty tests to give a hierarchy, without a Heywood case. But that cannot be easily said. It is impossible, from the correlations alone, to distinguish a real *g* from one imitated by a fortuitous coincidence of specifics. Even if *g* were a reality, a sample of persons equal in number to the tests could not give a hierarchy without a Heywood case, and their apparent *g* would be fortuitous.

Now the case of a test of pure *g* is on the border line of the Heywood cases. It is clear then that it will be suspect, as being probably only fortuitous, if the number of persons does not far exceed the number of tests.

7. *Singly conforming tests.*—There remains one other conceivable method of measuring *g* exactly,* by the use of certain tests which, when they are all present, destroy the hierarchy, although any one of them can enter the battery without marring it—" singly conforming " tests (Thomson, 1934*c*; and 1935*a*, 258–6). It will be remembered from the chapters on estimation that the reason factors cannot be measured exactly, but have to be estimated only, is that they outnumber the tests. Every new test which conforms to a hierarchy adds a new specific (unless it is pure *g*), and thus continues the excess of factors over tests. It can occur, however, that the correlation of two tests with each other breaks a hierarchy, although either of them alone conforms otherwise. Such a case occurs in the Brown-Stephenson battery, for example, one of whose correlation coefficients has to be suppressed before the hierarchy is acceptable.

In such a case, if the psychologist is prepared to accept

* By " exactly " is meant, with the same exactness as the test scores, without the additional indeterminacy due to an excess of factors over tests.

either test as a member of the battery, the erring correlation coefficient must be due to these two tests sharing some portion of their specifics with one another. If, as may happen (apart from error which we are supposing absent), their intercorrelation shows that they have only one specific factor between them, and differ only in their saturations, then they enable the estimate of g to be turned into accurate measurement. For example, consider the following matrix of correlations:

	1	2	3	4	5	6
1	.	·669	·592	·458	·335	·251
2	·669	.	·566	·438	·870	·240
3	·592	·566	.	·387	·283	·212
4	·458	·438	·387	.	·219	·164
5	·335	·870	·283	·219	.	·120
6	·251	·240	·212	·164	·120	.

This is a perfect hierarchy except for the correlation—

$$r_{25} = \cdot 870$$

Every tetrad-difference, which does not contain this correlation, is zero. If either Test 2 or Test 5 is removed from the battery, there remains a perfect hierarchy. If Test 5 is removed, we can calculate from the remaining battery the g saturations:

Test	1	2	3	4	6
g saturation	·837	·800	·707	·548	·300

If we remove Test 2 and restore Test 5, we get the following:

Test	1	3	4	5	6
g saturation	·837	·707	·548	·400	·300

From either hierarchy we can estimate g. The correlation of our estimates with "true g" will be—

$$\sqrt{\frac{S}{S+1}}$$

where $$S = \Sigma \frac{\text{saturation}^2}{1 - \text{saturation}^2}$$

and we find for the two hierarchies the g correlations of ·92 and ·90.

Suppose now that we had left both Tests 2 and 5 in the battery with which to estimate g, after calculating their g saturations from the two separate hierarchies, what influence would this have had upon the accuracy of our estimate ? It is of some interest actually to carry out this calculation by Aitken's method, using all the tests with the g saturations given above. A calculation keeping three places of decimals gives for the regression coefficients :

Test	1	2	3	4	5	6
Regression coefficient	·005	1·856	−·003	·001	−1·213	·002

which suggests (what would actually be the case if more decimals were retained throughout) that all the regression coefficients except those for Tests 2 and 5 vanish. If we calculate the multiple correlation of this battery with g, by finding the inner product of the g saturations with the above regression coefficients, we find that it is exactly unity.

The reason for this is that the correlation of Tests 2 and 5 is such as to show that their specifics are identical, the two tests differing only in their loadings. Their equations are—

$$z_2 = \cdot 8g + \sqrt{(1 - \cdot 8^2)}s_2$$
$$z_5 = \cdot 4g + \sqrt{(1 - \cdot 4^2)}s_5$$

If the whole of s_2 is identical with the whole of s_5, their intercorrelation should be—

$$\cdot 8 \times \cdot 4 + \sqrt{(1 - \cdot 8^2)(1 - \cdot 4^2)} = \cdot 870$$

and this is its experimental value.

We could, therefore, have seen at the beginning, if we had tested the above fact, that these two tests would make a perfect battery for measuring g. We have the simultaneous equations—

$$z_2 = \cdot 8g + \cdot 6s$$
$$z_5 = \cdot 4g + \cdot 917s$$

from which we can eliminate *s* by multiplying by—

$$\cdot 917 \quad \text{and} \quad -\cdot 600$$

respectively, numbers which are exactly in the ratio of the regression coefficients found above—

$$1\cdot 856 \quad \text{and} \quad -1\cdot 218.$$

In fact, we could have performed the regression calculation on these two tests alone, when it would have appeared as follows:

	1·000	·870	−1·000	.	·870
	·870	1·000	.	−1·000	·870
	·800	·400	.	.	1·200
(4·1135)		·2431	·8700	−1·000	·1131
		1·0000	3·5787	−4·1135	·4652
		−·2960	·8000	.	·5040
			1·8593	−1·2176	·6417

giving (more exactly) the same regression coefficients as before.

We see, therefore, that under certain hypothetical circumstances, a more exact estimate of *g* can be obtained from two of these "singly conforming" tests than the hierarchy with which they conform individually. Those circumstances are, that their correlation with one another (the correlation which breaks the hierarchy because it is too large) should either equal—

$$r_{ig}r_{jg} + \sqrt{(1 - r_{ig}^2)(1 - r_{jg}^2)}$$

or should approach this value.

It cannot in actual practice be expected to equal it, as in our artificial example. For we have disregarded errors, which are sure in some measure to be present. At what stage will the pair of singly conforming tests cease to be a better measure of *g* than the better of the two hierarchies made by deleting either the one or the other? If in our example the correlation ·870 of Tests 2 and 5 be imagined to sink little by little, the correlation of their estimate with *g* will sink from unity. The better of the two hierarchies gives a multiple correlation of ·922. When the

correlation r_{ss} has sunk from ·870 to ·847, these two singly conforming tests will give the same multiple correlation, ·922. If this defect from the full ·870 is due entirely to error, then a fall to ·847 corresponds to reliabilities of the two tests of the order of magnitude of ·98, if they are equally reliable. This is a very high reliability, seldom attained, so that in a case like our example quite a small admixture of error would make the singly conforming tests no better at estimating g than the hierarchy. We are here, however, neglecting the fact that error would also diminish the efficiency of the hierarchy. Nevertheless, the chance of finding a pair of singly conforming tests, highly reliable, and having no specifics except that which they share, seems small, as small as the chance of finding a test of pure g, perhaps. It might possibly turn out, however, that a matrix of several (say t) singly conforming tests would be practicable. Such a set would measure g exactly if among them they added only $t - 1$ new specifics to the hierarchy. Their saturations would be found by placing them one at a time in the hierarchy, and then their regression on g calculated by Aitken's method. The necessity for the hierarchy in the background, in all this, is clear: it is there to assure us that each singly conforming test is compatible with the definition of g, and to enable its g saturation to be calculated.

8. *The danger of " reifying " factors.*—The orthodox view of psychologists trained in the Spearman school is that g is, of all the factors of the mind, the most ubiquitous. " All abilities involve more or less g," Spearman has said, although in some the other factors are " so preponderant that, for most purposes, the g factor can be neglected." With this view, the present author has always agreed, provided that g is interpreted as a mathematical entity only, and judgment is suspended as to whether it is anything more than that.

The suggestion, however, that g is " mental energy," of which there is only a limited amount available, but available in any direction, and that the other factors are the neural machines, is one to be considered with caution. The word *energy* has a definite physical meaning. " Mental

energy " may convey the meaning that the energy spoken of is the same as physical energy, though devoted to mental uses. If that meaning is accepted, innumerable difficulties follow, not the least being the insoluble questions of the connexion of body and mind, and of freewill versus determinism. A less obscure difficulty is that there seems to be no easily conceivable way in which the " energy " of the whole brain can be used in any direction indifferently, except by the " neural engines " also all taking part. The energy of a neurone seems to reside in it, and the passage of a nerve impulse along a neurone seems to resemble rather the burning of a very rapid fuse, than the conduction of electricity, say, by a wire.

If " mental energy " does not mean physical energy at all, but is only a term coined by analogy to indicate that the mental phenomena take place " as if " there were such a thing as mental energy, these objections largely disappear. Even in physical or biological science, the things which are discussed and which appear to have a very real existence to the scientist, such as " energy," " electron," " neutron," " gene," are recognized by the really capable experimenter as being only manners of speech, easy ways of putting into comparatively concrete terms what are really very abstract ideas. With the bulk of those studying science there exists always the danger that this may be taken too literally, but this danger does not justify us in ceasing to use such terms. In the same way, if terms like " mental energy " prove to be useful, and can be kept in their proper place, they may be justified by their utility. The danger of " reifying " such terms, or such factors as g, v, etc., is, however, very great, as anyone realizes who reads the dissertations produced in such profusion by senior students using these new factorial methods.

CHAPTER XVI

" SIMPLE STRUCTURE "

1. *Simultaneous definition of common factors.*—In a sense, Thurstone's system of multiple common factors is a generalization of the original Spearman system which had only one. It recognizes that matrices of correlation coefficients are not usually reducible to rank 1, but that they are usually reducible to a low rank, and it replaces the analysis into one common factor and specifics by an analysis into several common factors and specifics, keeping the number of common factors at a minimum. It does not lay the great stress on the ubiquity and dominance of g which is found in the Spearman system. Indeed, in his latest analysis of a battery of fifty-seven tests (see Chapter XIX) Thurstone finds no general factor at all.

Spearman's system, having defined g as well as possible by an extended hierarchy, goes on then to definitions of the next most important factors, by similar means. It looks upon any complex matrix of correlations as being due to lesser hierarchies superimposed upon the g hierarchy. Moving in accordance with a very commonly held belief which almost certainly has some justification, it has sought and found " verbal " and " practical " factors to add to g, and is groping for some kind of character or emotional factor which would complete the main picture. " One at a time " has been its motto.

Moving along another route, Thurstone has endeavoured to define several factors by one matrix of correlations. Although the campaign of the Spearman school seems more practical, and was presumably, indeed, the only method open to pioneers, a student must be struck by the fact that the standard definition of g is made by a battery of tests (Brown and Stephenson, 1933) which is not really reducible to rank 1 until a large verbal factor has been

removed by mathematical means. Just as a battery to define g has to be purified either by the actual removal of tests or by the mathematical removal of factors before it is suitable as such a definition, so not every battery will define a group of common factors. Thurstone batteries, like Spearman's, have to be composed of selected tests, and purified if the selection is not complete. It is an obvious question, to ask whether different selections will lead to the same factors, or to incompatible sets.

2. *Incompatible sets conceivable.*—We saw in Chapter II that four tests, though they may give a matrix of correlation coefficients which can be reduced to rank 2, do not define two common factors, for the reduction can be made by many different sets of communalities; but that a fifth test, if its correlations still left rank 2 a possibility, fixed the communalities. These five tests, then, are a potential definition of two common factors, just as three tests defined one general factor; to speak more accurately, the five tests define a common-factor space of two dimensions within which the two common factors must lie, though they are not yet fully defined and may be rotated therein. The example used in Chapter II was the following:

	1	2	3	4	5*a*
1	.	·4	·4	·2	·5883
2	·4	.	·7	·3	·2852
3	·4	·7	.	·3	·2852
4	·2	·3	·3	.	·1480
5*a*	·5883	·2852	·2852	·1480	.

with the communalities—

·7	·7	·7	·1303	·5

Suppose, however, that when searching for a fifth test to add to the first four, which would give a matrix of rank 2, we had come across a test which gave the correlations shown in the fifth row and column here:

	1	2	3	4	5b
1	.	·4	·4	·2	·3530
2	·4	.	·7	·3	·3521
3	·4	·7	.	·3	·3521
4	·2	·3	·3	.	·2116
5b	·3530	·3521	·3521	·2116	.

This fifth test, which we shall call *5b*, also gives a matrix which is reducible to rank 2, but by communalities which in the first and fourth tests (especially the first) are incompatible with those fixed above, namely, by communalities—

·3 ·7 ·7 ·14 ·5

The factors given by an analysis of this matrix are not therefore the same as those formerly obtained, nor are they capable of being rotated into each other within the common-factor space. Indeed, the common-factor space of the matrix 1, 2, 3, 4, *5a* is quite different from that of the matrix 1, 2, 3, 4, *5b*. If we consider the matrix of all six tests, no matter what the correlation between tests *5a* and *5b* may be, we have a matrix which *cannot* be reduced to rank 2 :

	1	2	3	4	5a	5b
1	.	·4	·4	·2	·5883	·3530
2	·4	.	·7	·3	·2852	·3521
3	·4	·7	.	·3	·2852	·3521
4	·2	·3	·3	.	·1480	·2116
5a	·5883	·2852	·2852	·1480	.	*r*
5b	·3530	·3521	·3521	·2116	*r*	.

That is to say, it cannot be so reduced exactly, though if it were blurred by sampling error it might perhaps be "sufficiently" well represented by two common factors. But in time more exact experiment would bring to light the discrepancy. We have here to decide between three incompatible sets of factors :

(*a*) Those given by the matrix excluding *5b*.
(*b*) Those given by the matrix excluding *5a* ;
(*c*) Those given by the matrix of all six tests.

If this situation occurred, various circumstances might influence the decision, which set to accept. It might prove to be much easier to extend the matrix (*a*) than the matrix (*b*), by discovering tests to add to the battery without raising the reduced rank, in which case the pair of common factors corresponding to (*a*) would seem more useful and indeed more likely to be "real." It might prove practically impossible to extend either, in which case the more numerous common factors corresponding to (*c*) would probably be chosen. Throughout, the psychologist would be guided also by his psychological insight, or prejudices, in the matter.

8. *Heywood cases in multiple-factor batteries.*—Just as in a two-factor hierarchy we saw that a test might crop up whose loading with g exceeded unity (Heywood, 1931), thus demolishing the battery's utility as a definer of a general factor, so in the case of batteries with more than one common factor, tests may, conceivably, crop up which conform to the reduced rank of the pre-existing battery only if some test is given a communality exceeding unity; there may be as many of these in the battery as there are common factors, but this would involve some very high correlations. One case, however, can readily be introduced without arousing suspicions, e.g. the matrix—

	1	2	3	4	5*c*
1	.	·4	·4	·2	·619
2	·4	.	·7	·3	·185
3	·4	·7	.	·3	·185
4	·2	·3	·3	.	·094
5*c*	·619	·185	·185	·094	.

is reduced to rank 2 by communalities—

1·2 ·7 ·7 ·1294 ·32

which cannot correspond to any real pair of factors.

It is also logically possible, in a multiple-factor battery, to have as many tests with communalities of full unity, as there are factors in the battery, in which case these tests

would enable exact estimates of the factors to be made, without any indeterminacy. They would form a sub-battery to measure these factors, analogous to the logically conceivable (but not yet discovered) test of "pure *g*" in a hierarchy.

4. *Need for rotating the axes.*—Actually, batteries intended for analysis by the multiple-factor methods have not been built up from a small number of tests by adding others which preserve the reduced rank. Instead, experimenters have first assembled a number of tests which appeared to them to be likely to contain only, say, r common factors, factors which they have already suspected to exist and have tentatively named. They have then ascertained by using Thurstone's approximate communalities whether a reduced rank of r can be achieved as a sufficiently close approximation, by examining the residues after r factors have been "taken out." By analogy with Spearman's purification process, they might then remove any tests which were preventing this; but such purification has not been very usual though it seems just as justifiable here as in a hierarchy. Let us suppose that a battery, assembled because it appeared, psychologically, to contain r common factors, does give a matrix which can be reduced to rank r.

As was explained towards the end of Chapter II, the loadings given by the "centroid" process then include a number of negative values, and these the psychologist has difficulty in accepting in any large numbers. For it is hard for him to conceive of psychological factors which help in some tests and hinder in others, except in rare cases. The mathematician can then "rotate" the factor axes *within the common-factor space* (Thurstone's principle forbids him to go outside it) in search of a position which *will* satisfy the psychologist. One way of doing this has already been sketched in Chapter II, Section 8. It has been used with excellent effect by W. P. Alexander (Alexander, 1935), but involves assuming (*a*) that the communality of a certain test is entirely due to one factor; (*b*) that the communality of a second test is entirely due to this factor and one other, (*c*), and so on for $r - 1$ tests,

where r is the number of factors. The criterion of success with this method is to see whether, when these assumptions are made, negative loadings disappear; and whether the consequent loadings of those tests about which no assumptions are made are compatible with the psychologist's *psychological* analysis of them. It cannot be too emphatically pointed out that the first factors which emerge from the " centroid " process and the minimum-rank principle need not have psychological significance as *unitary* primary traits. It is only after rotation to a suitable position that this can be expected.

5. *Agreement of mathematics and psychology.*—It becomes increasingly clear that the whole process is one by which a definition of the primary factors is arrived at by satisfying simultaneously certain mathematical principles and certain psychological intuitions. When these two sides of the process click into agreement, the worker has a sense of having made a definite step forward. The two support one another. Obviously the goal to be hoped for along this line of advance will be the discovery of some mathematical process which always leads to a unique set of factors mainly acceptable to the psychologist. If such could be discovered and found to produce a few factors over and above those recognized as already known by other means, the new factors would stand a good chance of acceptance on the strength of their mathematical descent only. And no doubt the psychologist would be prepared to make a few concessions and changes in his previous ideas to fit in with any mathematical scheme which already gave much satisfaction and was objective and unique in its results.

We have, it is true, already seen reason to doubt whether *any* process can always lead to one universal set of factors. Different batteries with some tests in common may lead to incompatible sets of factors; selection will change factors; and so on. But let us suppose that these difficulties are overcome somehow. Perhaps incompatible batteries though logically conceivable do not actually occur. Perhaps we may outflank the selection difficulty by defining our population arbitrarily. Let us assume that the principle of employing the minimal reduced rank

(criticized in Chapter VIII) has, nevertheless, justified itself. We have arrived at a common-factor space, the dimensions of which may not, by the principles we have adopted, be altered. We need, however, to complete our scheme, some objective means of rotating the factor axes in this common-factor space to a unique final position, and we do not want to do this by the somewhat crude method already mentioned of assuming the absence of factors in certain tests. It is here that Thurstone's notion of "simple structure" is offered as a solution (*Vectors*, Chapters 6–8). This idea is that the axes are to be rotated until as many as possible of them are at right angles to as many as possible of the original test vectors; and that the battery is not suitable for defining factors unless such a rotation is uniquely possible, a rotation which will leave every factor axis at right angles to at least as many tests as there are factors, and every test at right angles to at least one factor.

When the vectors of a test and a factor are at right angles, the loading of the factor in that test is zero. Thurstone's "simple structure" is therefore indicated by a large number of zeros in the matrix of loadings, so large that there will be only one position of the axes (if any) which satisfies the requirement. His search, be it repeated, is for a set of conditions which will make the solution unique. We have seen him approaching this goal by stages. Unless the battery is large, so that—

$$n \geqslant \frac{(2r+1)+\sqrt{(8r+1)}}{2}$$

(see Chapter II, Section 9), the communalities are not unique. Even when the battery is large enough, the axes representing factors may be rotated to positions among which there is no one specially marked out. Then comes the demand that there be this large number of zero loadings. Most batteries of tests will not allow this demand to be satisfied, but with some it can just be attained. Only these last, it is Thurstone's conviction, are suitable for defining primary factors, and it is his faith that the factors

thus mathematically defined will be found to be acceptable as psychologically separable unitary traits.

6. *An example of six tests of rank 3.*—To make our remarks more definite and concrete, let us suppose that we have a battery of six tests whose matrix of correlations can be reduced to rank 3. The number of tests fulfils the inequality requirement, and this set of communalities is therefore unique. The matrix of loadings given by the "centroid" system contains at first negative quantities. Thus from the correlations:

	1	2	3	4	5	6
1	.	·525	·000	·000	·448	·000
2	·525	.	·098	·306	·349	·000
3	·000	·098	.	·133	·314	·504
4	·000	·306	·133	.	·000	·000
5	·448	·349	·314	·000	.	·307
6	·000	·000	·504	·000	·307	.

with the communalities—

·674	·634	·558	·415	·490	·493

we get by the "centroid" process the matrix of loadings:

	I	*II*	*III*
1	·542	·612	·074
2	·629	·342	—·348
3	·529	—·492	·191
4	·281	—·182	—·550
5	·628	·148	·274
6	·429	—·424	·359

It is the factor axes indicated by these loadings that Thurstone wishes to rotate until there are no negative loadings and enough zero loadings to make the position uniquely defined. For this last purpose he finds, empirically, that it is necessary to require—

(*a*) At least one zero loading in each row ;

(*b*) At least as many zero loadings in each column as there are columns (here three); and

(*c*) At least as many *XO* or *OX* entries in each pair of columns as there are columns. By an *XO* entry is meant a loading in the one column opposite a zero in the other.

Now, these requirements cannot generally be met by a matrix of loadings. It will in general be impossible to rotate the axes until every axis is at right angles to *r* test vectors. The above example has, however, been constructed so that this can be done. The loadings can be rotated into the form:

	I	*II*	*III*
1	.	.	·821
2	.	·475	·639
3	·718	·206	.
4	.	·644	.
5	·438	.	·546
6	·702	.	.

which satisfies the three conditions enumerated, and where, moreover, the factors are still orthogonal.

Before turning to the consideration of ways of testing whether such "simple structure" can be reached in a given case, and ways of reaching it when it is possible, it is advisable to dwell for a while on the significance of Thurstone's three requirements, for it will be by bearing them in mind that the experimentalist can hope to build up a matrix permitting their fulfilment.

"At least one zero loading in each row." This means that no test may contain all the common factors. In making up the battery, then, the experimenter, with some idea in his mind as to what the factors are, will endeavour to ensure that they are not all present in any one test. This would, for example, exclude from a Thurstone battery any very mixed group test, or a mixed test like the Binet-Simon which is itself a whole battery of varied items.

"At least as many zeros in each column as there are columns," that is, as there are common factors. This

means that in a Thurstone battery no factor may be general, but must be missing in several tests. This would, for example, require that several of the tests have zero saturation with Spearman's factor *g*, a somewhat difficult requirement to meet, one would think, except approximately.

The requirement as to the number of *XO* or *OX* entries is intended to ensure that the tests are qualitatively distinct from one another. For example, if the entry ·438 in the above matrix of loadings were moved up from Test 5 to Test 4, then Tests 1 and 5 would each have two zero loadings in Factors I and II and would differ only in that their saturations with Factor III are different, namely, ·821 and ·546. When the rule about *XO* entries is fulfilled, all the tests differ qualitatively, as it were, and not merely quantitatively.

7. *Devices for finding simple structure.*—Whatever mathematical devices may be discovered for rotating the matrix of loadings from the first form to that of " simple structure " (when the latter can be attained) it seems probable that a large part will be played by the intuition of the psychologist in knowing which cells of the matrix are likely to be reducible to blanks. Thus in the above example, if the psychologist had a previous inkling as to the nature of the three factors, and that Tests 1, 4, and 6 each contained only one of them, he could very rapidly have calculated the " simple structure " form of the loadings. One feels here the danger of a certain amount of self-deception. A mathematical method and psychological intuitions are being brought into agreement by picking those tests to form the battery which permit that agreement. The agreement, when it is arrived at, is perhaps liable to impress the psychologist too much, and make him feel that the existence of the factors, in terms of which he has been thinking while picking the tests, has been completely proved by the possibility of making a " simple structure." It is very hard to say just what has been proved in that case : and in any case experiment has not yet produced many batteries which clearly exhibit this phenomenon. Undoubtedly, however, if more such bat-

teries were produced (in the same way as Brown and Stephenson have produced the hierarchical battery of nineteen tests), and if the resulting "simple structure" factors were compatible with one another and in fair agreement with psychological intuition, there would be formed an apparatus for defining factors which would have considerable influence on the progress of psychology.

Among the devices which Thurstone has used to aid in finding the proper factors is that of removing from the battery all those tests which appear, psychologically, to contain a certain factor, and then checking whether the reduced rank of the remaining battery has fallen by one from its former value.

Another device is to search for "clusters" among the correlation coefficients after the latter have been "corrected for communality," as it is called, though the term "corrected" is not a good one here. By "correcting" a correlation coefficient for communality is meant dividing it by the square root of each of the communalities of the two tests concerned. The result is the correlation which would ensue if the specifics were abolished. It is, of course, a higher correlation, for specifics dilute the resemblance between tests: and in the case of two tests which were identical in both loadings and factors, except for their specific, this "corrected" correlation would be unity.

In the case of our small example, the correlations "corrected" for communality are:

	1	2	3	4	5	6
1	1·000	·803	·000	·000	·780	·000
2	·803	1·000	·165	·597	·626	·000
3	·000	·165	1·000	·276	·601	·961
4	·000	·597	·276	1·000	·000	·000
5	·780	·626	·601	·000	1·000	·626
6	·000	·000	·961	·000	·626	1·000

Of course, in a small artificial example with only six tests one cannot expect to be able to talk of "clusters" of similar tests—similar except for their specifics. But the high value ·961 catches the eye, and indicates that Tests

3 and 6 are qualitatively very much alike except for their specifics. Also the value ·808 draws attention to the similarity of Tests 1 and 2, which we note to be almost entirely uncorrelated with the previous "cluster" of Tests 3 and 6. This all suggests that our rotated matrix of marks may begin by having loadings as shown here:

	A	B	C
1	O	X	?
2	O	X	?
3	X	O	?
4			
5			
6	X	O	?

the first pair being mainly composed of a factor here called A, the second pair mainly of B, and the loadings marked with a query being small if not zero. This agrees with our rotated loadings, A being Factor I and B being Factor III.

Thurstone's final device for obtaining the position of simple structure is to form the sum, in each column, of the quantities—

$$w = \frac{1}{\text{loading}^2 + \cdot 01}$$

This expression, it will be observed, becomes large when the loading is zero. It would indeed become infinite (which would be inconvenient) were not the small quantity ·01 added, which makes its upper limit 100. In our example, the sum of these quantities for the unrotated form of the first-factor loadings was 27·97, but for the rotated form 308·85.

Test	Unrotated Loadings		Rotated Loadings	
	I	w	I	w
1	·542	3·19	.	100·00
2	·629	2·47	.	100·00
3	·529	3·45	·718	1·90
4	·281	11·24	.	100·00
5	·628	2·47	·438	4·96
6	·429	5·15	·702	1·99
		27·97		308·85

Thurstone searches, by a directed form of trial and error,* for the loadings which make Σw a maximum (see *Vectors*, pages 182–5). This will not necessarily be the position with most zeros, but it is likely to be near it. Of course, throughout the trials the loadings have to fulfil certain conditions, and have to give the same correlations between the tests however they may be changed. One of the conditions which we have hitherto imposed, however, Thurstone relaxes, namely that the factors be orthogonal.

8. *Oblique factors.*—It is natural to desire factors to be orthogonal, that is independent, uncorrelated with one another. In describing a man, or an occupation, by means of factors it would be both confusing and uneconomical to use factors which, as it were, overlapped. Yet in situations where more familiar entities are dealt with we do not hesitate to use correlated measures in describing a man. For instance, we give a man's height and weight, although these are correlated quantities. But if we are going to allow factors to be used which are as highly correlated as this, there seems no reason to use *fictitious* factors at all; we might as well use certain tests just as they stand, as was suggested in the opening pages of Chapter I.

* We learn from his latest work (see our Chapter XIX) that he has returned to the device of rotating the factors pair by pair graphically.

It does not seem, however, that Thurstone wishes to use factors which are really correlated in the whole population, but that he recognizes that they are unlikely in that case to be *exactly* uncorrelated in the experimental sample. He is therefore willing to let the right angles between the factors, as expressed by the loadings, sag away from strict rectangularity if that will give him the number of zeros required by simple structure. The same argument would also lead one to be lenient in accepting small loadings, negative or positive, as equivalent to zero.

As soon as we allow oblique factors it is necessary to make a distinction between what Holzinger calls pattern and structure.

9. *Pattern and structure.*—So long as the factors are orthogonal, the loadings in the matrix of loadings are also the correlations between the factor and the tests, but this ceases to be the case when the factors are correlated. The word " loading " continues to be used for the coefficients such as l, m, and n in equations like—

$$z = l\alpha + m\beta + n\gamma$$

and the matrix or table of these is called a *pattern*, while the matrix of correlations between tests and factors is called a *structure*. Thus of the two matrices on page 182 (Chapter XI, Section 6), the upper one is both a pattern and a structure, for the factors are orthogonal, whereas the lower one is a structure only. From the upper table we can say that—

$$z_1 = \cdot 70f_1 + \cdot 40f_2 + \cdot 59s_1$$

using the correlations of the factors with Test 1 as coefficients in a linear equation for that test score. But we cannot say from the lower table that—

$$z_1' = \cdot 51f_1 + \cdot 25f_2 + \cdot 40s_1$$

The correlations here cannot serve as coefficients. There is a simple algebraic connexion between structure and pattern which is deduced in the Appendix, paragraph 19. Since structure and pattern deviate from one another with oblique factors, and since Thurstone is prepared to admit factors which, in the experimental sample at any rate, are

somewhat oblique, the question may arise whether the zeros he demands are to be in the pattern of loadings or in the structure of correlations. Holzinger in his *Manual* (Holzinger, 1937, 68 and 74) discusses both possibilities. From the name " simple *structure* " it would seem that essentially it is the structure and not the pattern that Thurstone has in mind. But with slightly oblique factors they will not differ much.

CHAPTER XVII

LIMITS TO THE EXTENT OF FACTORS

1. *Boundary conditions in general.*—Before we discuss further the question whether a given set of common-factor loadings can be rotated into "simple structure" it is desirable to consider a wider problem, in itself quite unconnected with Thurstone's particular theory of factors; the problem, namely, of drawing conclusions from correlation coefficients as to what there is in common between tests, or other variates. From one correlation coefficient, if it is significant in proportion to its standard error, it is natural to assume that the variates share some causal factor, though that factor may be a very abstract thing. But the circumstance that the correlation is not perfect shows that other causal factors too are at work. These may dilute the correlation in various ways. Some cause may be influencing the variate (1) but not the variate (2). Or vice versa some cause may be influencing (2) but not (1). Or both these things may be happening. Or some cause may be helping the one variate, and hindering the other. In any case, however, if the two variates are expressed as weighted sums of uncorrelated factors—

$$z_1 = l_1a_1 + l_2a_2 + l_3a_3 + \dots$$
$$z_2 = m_1b_1 + m_2b_2 + m_3b_3 + \dots$$

one at least of the factors *a* must be identical with one at least of the factors *b*, in order that any correlation may result.

If we next consider three tests and low correlations (up to ·5), we find great elasticity in the possible explanations.* Suppose all three correlations equal ·5. We have, then, among innumerable possibilities, two extreme forms of

* Brown and Thomson, page 142; Thomson, 1919*b*, Appendix J. R. Thompson.

explanation possible, one with only one general factor, the other with no general factor—

$$\left.\begin{aligned} z_1 &= \cdot707a + \cdot707s_1 \\ z_2 &= \cdot707a + \cdot707s_2 \\ z_3 &= \cdot707a + \cdot707s_3 \end{aligned}\right\} \text{one general factor}$$

or

$$\left.\begin{aligned} z_1 &= \cdot707b + \cdot707c \\ z_2 &= \phantom{\cdot707b + {}} \cdot707c + \cdot707d \\ z_3 &= \cdot707b \phantom{{} + \cdot707c} + \cdot707d \end{aligned}\right\} \text{no general factor}$$

So long as the correlations do not average more than ·5,* they can (usually) be imitated without a general factor, although one can be used if desired. That is, they can be imitated either by a three-factor—if we may so designate a factor running through three tests—or (usually) by two-factors running through only two tests, though in certain cases this may prove impossible, especially if the average correlation is not far below ·5.

As soon, however, as the average correlation rises above ·5,† some use *must* be made of a three-factor general to all three tests, as the reader can readily convince himself by trial. In the above example, if we wish to increase the correlation of Tests 1 and 2 while using the second form of equations, we see that since we have exhausted all the variance on the factors *b*, *c*, and *d*, we can do so only by using either *b* in Test 2, or *d* in Test 1, and thus making it into a three-factor.

2. *The average correlation rule* (Thomson, 1936*b*). When we have more tests, say *n*, then we can usually do without an *n*-factor (or general factor) so long as the average correlation does not exceed $(n - 2)/(n - 1)$.‡ Again, of course, an *n*-factor may be used if desired, but its use is not usually compulsory, as it certainly is in some measure as soon as

* This is an approximate condition. For an exact form, see the Mathematical Appendix, paragraph 20. See also later in this chapter.

† See previous footnote.

‡ Approximate condition, see previous footnote, and consult Appendix.

the average correlation rises past this point. Further, if the average correlation is still lower, we can in turn, as a rule, dispense with $(n - 1)$-factors as soon as the average sinks below $(n - 3)/(n - 1)$, and with factors of less extent as it sinks still further. To know approximately what is the least-extensive kind of factors we can manage with, we have to see where the average correlation fits in, in the series of fractions—

$$\frac{1}{n-1}\quad\frac{2}{n-1}\quad\frac{3}{n-1}, \ldots \frac{n-3}{n-1}\quad\frac{n-2}{n-1}$$

As soon as the average correlation rises *past* $(n-p)/(n-1)$, we can no longer have $(p - 1)$ zeros in every column of the matrix of loadings. Usually (though not necessarily) we can manage to have $(p - 1)$ zeros at or below that point.

The reason for this rule can be appreciated if we reflect that the highest possible correlations we can get with a given number of zero loadings will be reached by abolishing all factors of less extent. For example, with two-factors only, the highest possible correlations between five tests will be obtained by a pattern of loadings like this:

$$\begin{matrix} X & X & X & X & 0 & 0 & 0 & 0 & 0 & 0 \\ X & 0 & 0 & 0 & X & X & X & 0 & 0 & 0 \\ 0 & X & 0 & 0 & X & 0 & 0 & X & X & 0 \\ 0 & 0 & X & 0 & 0 & X & 0 & X & 0 & X \\ 0 & 0 & 0 & X & 0 & 0 & X & 0 & X & X \end{matrix}$$

If there are to be no specifics, and if we take the case where all the correlations are alike (which is in fact the maximum correlation possible), we see that the square of every loading must be $1/4$, or in general $1/(n - 1)$. Each correlation will therefore be equal to $1/4$ or $1/(n - 1)$. In the series of fractions—

$$\frac{1}{4}\quad\frac{2}{4}\quad\frac{3}{4}$$

the average correlation just reaches the first, which can be considered as $\frac{n-p}{n-1}$, n being 5 and p being 4. And $p - 1$

or three zeros are just possible in each column of loadings.

Again, consider five tests in which we use only three-factors. The maximum correlation is given by a pattern just like the last one, except that the noughts and crosses have to change places. Since there are six loadings, the square of every loading must be 1/6, and the pattern shows that every correlation is three times this, or 1/2. The average correlation, therefore, now reaches the next of the above fractions—

$$\frac{1}{4} \quad \frac{2}{4} \quad \frac{3}{4}$$

and when

$$\frac{n-p}{n-1} = \frac{2}{4}$$

we have $p = 3$; and $p - 1$ or two zeros are just possible per column (represented by the crosses in the former diagram), as we know is true from the way in which we made the correlations.

It should be noted that the rule works with certainty only in one direction. What it asserts to be impossible, is impossible. But when it does not say that a given number of zero loadings per column is impossible, it is not certain to be possible. The rule is necessary, but not sufficient. Usually, however, it is a fairly safe guide, and when it does not say the zeros are impossible, they can generally be nearly if not quite reached, with the greater ease, of course, the more the average correlation falls below the critical value.

It should also be re-emphasized that these considerations have, so far, nothing to do with Thurstone's theory. In terms of our geometrical analogy, we are here considering the whole space (not merely a common-factor space) and asking whether orthogonal axes can be found each of which is at right angles to some of the test vectors. We are at liberty to take as many axes as we like, extending the dimensions of our space as we please.

As an example, consider the set of correlations used in the last chapter:

	1	2	3	4	5	6
1	.	·525	·000	·000	·448	·000
2	·525	.	·098	·306	·349	·000
3	·000	·098	.	·133	·314	·504
4	·000	·306	·133	.	·000	·000
5	·448	·349	·314	·000	.	·307
6	·000	·000	·504	·000	·307	.

The average correlation is ·199, and n, the number of tests, is 6. The series of critical fractions is therefore—

$$\frac{1}{5} \quad \frac{2}{5} \quad \frac{3}{5} \quad \frac{4}{5}$$

and the average correlation falls just short of the first one, for which, since $n - p = 1$, $p = 5$. This leaves open the possibility that we can use factors which have $p - 1$ or four zeros in each column of loadings, that is, that we can manage with two-factors each linking only two tests. But as ·199 is so near to 1/5, and as the correlations are far from being all alike, we may expect to find this difficult or even not quite possible. Trial shows that we can *nearly*, but not quite, manage with two-factors. The following set of loadings, for example, while not perhaps the nearest approach to success, comes fairly close :

Factor	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*
Test									
1	·734	·679	.	.	.	.	.	.	.
2	·658	.	·300	·318	·613	.	.	.	.
3	.	.	·301	.	.	·278	·606	·682	.
4	.	.	.	·895	.	·446	.	.	.
5	.	·607	.	.	·504	.	·477	.	·387
6	.	.	.	.	.	.	.	·682	·732

giving correlations :

	1	2	3	4	5	6
1	.	·483	·000	·000	·412	·000
2	·483	.	·090	·285	·309	·000
3	·000	·090	.	·124	·289	·465
4	·000	·285	·124	.	·000	·000
5	·412	·309	·289	·000	.	·283
6	·000	·000	·465	·000	·283	.

which average ·188 instead of ·199.

8. *The latent-root rule* (Thompson, 1929; Black, 1929; Thomson, 1936*b*; Ledermann, 1936).—A more scientific rule for ascertaining how "extensive" * the factors must be to explain the correlations is based upon the calculation of the largest "latent root" of the matrix of correlations. The exact calculation of the largest latent root is a very troublesome business, but luckily there are approximations. We have already met the term "latent root," in passing, in connexion with Hotelling's process.†

If the largest latent root lies between the integers s and $(s + 1)$, then s-factors are certainly unable to imitate the correlations. Like the previous rule, this one is "necessary," but not "sufficient." It assures us that s-factors are inadequate, but it does not assure us that $(s + 1)$-factors are adequate, though they usually are if the latent root is not too near $s + 1$.

The easiest approximation to the largest latent root is, when the correlations are positive—

$$\frac{\text{Sum of the whole matrix, including diagonal elements}}{n}$$

In the case of the above example the whole matrix, including unities in the diagonal elements, sums to 11·972, so that the approximate largest latent root is 1·995, which leaves it just barely possible that two-factors will suffice. As we know by trial, they just won't.

A better approximation is—

$$\frac{\text{Sum of the squares of the column totals}}{\text{Sum of the whole matrix}}$$

the diagonal elements being included for both numerator and denominator. (This quantity is, in fact, the sum of the squares of the first-factor loadings in Thurstone's "centroid" process.)

* Meaning by an "extensive" factor one which has loadings in many tests. Thus a two-factor is less "extensive" than a three-factor, and so on.

† See Chapter V, Section 4.

In our example we have:

	1·000	·525	·000	·000	·448	·000
	·525	1·000	·098	·306	·349	·000
	·000	·098	1·000	·133	·315	·504
	·000	·306	·133	1·000	·000	·000
	·448	·349	·315	·000	1·000	·308
	·000	·000	·504	·000	·308	1·000
Totals	1·973	2·278	2·050	1·439	2·420	1·812 = 11·972
Squares	3·893	5·189	4·203	2·071	5·856	3·283 = 24·495

$$\text{Approximate largest latent root } \frac{24{\cdot}495}{11{\cdot}972} = 2{\cdot}046 \text{ *}$$

This time the better approximation definitely cuts out the possibility that two-factors will suffice.

4. *Application to the common-factor space.*—All of the above applies to factors in general, and the calculations are carried out with unity in each diagonal cell. To apply these rules to the problem of the attainability of "simple structure," we have to adapt them to the common-factor space. For this purpose they must be applied either to the matrix with correlations "corrected" for communality (the best plan), or with certain modifications to the matrix with communalities in the diagonal. The correlations "corrected" for communality are given on page 252 of Chapter XVI. The average of the correlation coefficients is ·362. In the series of fractions with denominator $(n - 1)$—

$$\frac{1}{5} \quad \frac{2}{5} \quad \frac{3}{5} \quad \frac{4}{5}$$

this value ·362 is below 2/5, or $(n - p)/(n - 1)$ where $n = 6$ tests. We see, therefore, that $p = 4$, and that the possibility of having $p - 1$, or three zeros in every column, is not denied. This is in agreement with the analysis (an orthogonal "simple structure") arrived at in Chapter XVI, page 250.

The first approximation to the largest latent root of the matrix with correlations "corrected" for communality

* The exact value to three places of decimals calculated by the method given in Aitken, 1937*b*, 284 *ff*., is 2·086.

and with unity in each diagonal cell (Chapter XVI, page 252), gives—

$$\frac{\text{Sum of whole matrix}}{n} = 2{\cdot}812$$

and as this is less than 3, three zeros are still possible in each column. The more accurate approximation to the root—

$$\frac{\text{Sum of the squares of the column totals}}{\text{Sum of the whole matrix}} = \frac{49{\cdot}2718}{16{\cdot}870} = 2{\cdot}92$$

shows by its nearness to 3 that three zeros, if they are possible (and we know they are), must just barely be possible.*

Instead of applying the latent-root test to the matrix corrected for communality, we can apply it to the matrix of ordinary correlations, with the communalities in the diagonal cells, but with the following change. Instead of comparing the latent root with the series of integers 1, 2, 3 . . . we have to compare it with the sum of 1, 2, 3 . . . communalities, taking these in their order of magnitude, largest first (Ledermann). We shall illustrate this on the same example. The matrix of ordinary correlations, with communalities, is:

	·674	·525	·000	·000	·448	·000	
	·525	·634	·098	·306	·349	·000	
	·000	·098	·558	·133	·314	·504	
	·000	·306	·133	·415	·000	·000	
	·448	·349	·314	·000	·490	·000	
	·000	·000	·504	·000	·000	·493	
Sums	1·647	1·912	1·607	·854	1·601	·997	= 8·618
Squares	2·713	3·656	2·582	·729	2·563	·994	= 13·237

$$\text{Approximate largest root } \frac{13{\cdot}237}{8{\cdot}618} = 1{\cdot}536$$

* Exact root is 2·954. It is tempting to surmise that Thurstone's search for *unique* orthogonal simple structure is really a search for a matrix, corrected for communality, with an *integral* largest root, equal to r; but it must be remembered that the criterion though necessary is not sufficient when the number of factors is restricted to r,

The communalities arranged in order of magnitude and summed are:

	1	2	3	4	5	6
	·674	·634	·558	·493	·490	·415
Continued sum	·674	1·308	1·866	2·359	2·849	3·264

The latent root 1·536 is larger than the second of these but less than the third, so the possibility of three zeros per column is left open, in agreement with the former tests and with the known facts. It would seem from the present writer's experience, however, that the test applied to the ordinary matrix in this way does not always agree exactly with that applied to the matrix with correlations corrected for communality, and that the latter is more accurate.

5. *A more stringent test.*—The above tests only refer to the possibility of obtaining the required number of zero loadings with *orthogonal* factors—"orthogonal simple structure." Even when orthogonal simple structure cannot be reached, it may be possible to attain simple structure with *oblique* factors.

Moreover, the approximations used for the largest latent root above are only valid, in general, when all the correlations are positive. In view of the fact, however, that few psychological correlations are negative this is not a great difficulty.

Further, while these tests show definitely when orthogonal simple structure cannot be attained, it does not follow with certainty that it can actually be reached when the tests are satisfied, though it usually can.

An exact criterion has been given (Ledermann, 1936), and is described in the Appendix, which avoids all the above defects. It requires at present, however, a prohibitive amount of calculation.

In general, simple structure will be attainable with a battery of tests only when the battery has been picked with that end in view. There is a certain incompatibility about Thurstone's demands which makes their fulfilment only possible in special circumstances. He wants as few common factors as possible to explain the correlations; but he wants these common factors to have no loadings

in a large number of the tests. This is rather like wanting to run a school with as few teachers as possible, but each teacher to have a large number of free periods. If we begin by reducing the number of common factors to its minimum (as Thurstone does), we will generally find that the second requirement cannot be fulfilled. It can, however, be fulfilled in some cases, and it is exactly these cases which Thurstone relies on to define his primary factors. It is his faith that factors found in this mathematical way will turn out to be acceptable to the psychologist as psychological entities.

CHAPTER XVIII

THE SAMPLING OF BONDS

1. *Brief statement of views.*—The purpose of this chapter is to give an account of the author's own views as to the meaning of "mental factors." This can perhaps be done most clearly by first expressing them somewhat emphatically and crudely, and afterwards adding the details and conditions which a consideration of all the facts demands. In brief, then, the author's attitude is that he does not believe in factors if any degree of real existence is attributed to them; but that, of course, he recognizes that any set of correlated human abilities can always be described mathematically by a number of variables or "factors," *and that in many ways*, among which no doubt some will be more useful or more elegant or more sparing of unnecessary hypotheses. But the mind is very much more complex, and also very much more an integrated whole, than any naïve interpretation of any one mathematical analysis might lead a reader to suppose. Far from being divided up into "unitary factors," the mind is a rich, comparatively undifferentiated complex of innumerable influences—on the physiological side an intricate network of possibilities of intercommunication. Factors are fluid descriptive mathematical coefficients, changing both with the tests used and with the sample of persons, unless we take refuge in sheer definition based upon psychological judgment, which definition would have to specify the particular battery of tests, and the sample of persons, as well as the method of analysis, in order to fix any factor. Two experimental observations are at the bottom of all the work on factors, the one that most correlations between human performances are positive, the other that square tables of correlation coefficients in the realm of mental measurement tend to be reducible to a low rank by suitable diagonal elements. The first of these (i.e. the predomi-

nance of positive correlations) appears to be partly a mathematical necessity, and partly due to survival value and natural selection. The second (i.e. the tendency to low rank) is a mathematical necessity if the causal background of the abilities which are correlated is comparatively without structure, so that *any* sample of it can occur in an ability. This enables one to say that the mind works *as if* it were composed of a smallish number of common faculties *and a host of specific abilities*; but the phenomenon really arises from the fact that the mind is, compared with the body, so Protean and plastic, so lacking in separate and specialized organs.

2. *Negative and positive correlations.**—The great majority of correlation coefficients reported in both biometric and psychological work are positive. This almost certainly represents an actual fact, namely that desirable qualities in mankind tend to be positively correlated; for though reported correlations may be selected by the unconscious prejudices of experimenters, who are usually on the look-out for things which correlate positively, yet as those who have tried know, it is really very difficult to discover negative correlations between mental tests. Besides, even in imagination we cannot make a race of beings with predominantly negative correlations. A number of lists of the same persons in order of merit can be all very like one another, can indeed all be identical, but they cannot all be the opposite of one another. If Lists a and b are the inverse of one another, List c, if it is negatively correlated with a, will be positively correlated with b. Among a number n of variates, it is logically possible to have a square table of correlation coefficients each equal to unity; that is, an average correlation of unity. But the farthest the average correlation can be pushed in the negative direction is $-1/(n-1)$. That is, if n is large, the average correlation can range from $+1$ to only very little below zero. Even Mother Nature, then, by natural selection or by any other means, could not endow man

* This section refers to correlations between *tests*. The greater frequency of negative correlations between *persons* has already been discussed in Chapter XIII, Section 8.

with abilities which showed both many and large negative correlations. If they were many, they would have to be very small; if they were large, they would have to be very few.

Natural selection has probably tended, on the whole, to favour positive correlations within the species.* In the case of some physical organs it is obvious that a high positive correlation is essential to survival value—for example, between right and left leg, or between legs and arms. In these cases of actual paired organs, however, it is doubtless more than a mere figure of speech to speak of a common factor as the cause. Between organs not simply related to one another, as say eyes and nose, natural selection, if it tended towards negative correlation, would probably split the genus or species into two, one relying mainly on eyesight, the other mainly on smell. Within the one species, since it is mathematically easier to make positive than negative correlations, it seems likely that the former would largely predominate. To say that this was *due to*

* An important kind of natural selection is the selection of one sex by the other in mating. Dr. Bronson Price (1936) has pointed out that positive cross-correlation in parents will produce positive correlation in the offspring Price further shows that this positive cross-correlation in the parents will result if the mating is highly homogamous for total or average goodness in the traits, a conclusion which, it may be remarked here, can be easily seen by using the pooling square described in our Chapter VI. Price concludes: " The intercorrelations which *g* has been presumed to illumine are seen primarily as consequences of the social and therefore marital importance which has attached to the abilities concerned." Price in his argument makes use of formulæ from Sewall Wright (1921). M. S. Bartlett, in a note on Price's paper (Bartlett, 1937*b*), develops his argument more generally, also using Wright's formulæ, and says: " Price contrasts the idea of elementary genetic components with factor theories. . . . It should, however, be pointed out that a statistical interpretation of such current theories can be and has been advocated. Thomson has, for example, shown . . .", and here follows a brief outline of the sampling theory. " On the basis of Thomson's theory," Bartlett adds, " I have pointed out (Bartlett, 1937*a*) that general and specific abilities may naturally be defined in terms of these components, and that while some statistical interpretation of these major factors seems almost inevitable, this may not in itself render their conception invalid or useless."

a general factor would be to hypostatize a very complex and abstract cause. To use a general factor in giving a description of these variates is legitimate enough, but is, of course, nothing more than another way of saying that the correlations are mainly positive—if, as is the case, most people mean by a general factor one which *helps* in every case, not an interference factor which sometimes helps and sometimes hinders.

3. *Low reduced rank.*—It is, however, on the tendency to a low reduced rank in matrices of mental correlations that the theory of factors is mainly built. It has very much impressed people to find that mental correlations can be so closely imitated by a fairly small number of common factors. Ignoring the host of specific factors to which this view commits them, they have concluded that the agreement was so remarkable that there must be something in it. There is; but it is almost the opposite of what they think. Instead of showing that the mind has a definite structure, being composed of a few factors which work through innumerable specific machines, the low rank shows that the mind has hardly any structure. If the early belief that the reduced rank was in all cases *one* had been confirmed, that would indeed have shown that the mind had no structure at all but was completely undifferentiated. It is the *departures* from rank 1 which indicate structure, and it is a significant fact that a general tendency is noticeable in experimental reports to the effect that batteries do not permit of being explained by as small a number of factors in adults as in children, probably because in adults education and vocation have imposed a structure on the mind which is absent in the young.*

By saying that the mind has little structure, nothing derogatory is meant. The mind of man, and his brain, too, are marvellous and wonderful. All that is meant by the absence of structure is the absence of any fixed or strong linkages among the elements (if the word may for a moment be used without implications) of the mind, so that any sample whatever of those elements or components can be assembled in the activity called for by a " test."

* See also Anastasi, 1936.

Not that there is any necessity to suppose that the mind is composed of separate and atomic elements. It is possibly a continuum, its elements if any being more like the molecules of a dissolved crystalline substance than like grains of sand. The only reason for using the word " elements " is that it is difficult, if not impossible, to speak of the different parts of the mind without assuming some " items " in terms of which to think. For concreteness it is convenient to identify the elements, on the mental side, with something of the nature of Thorndike's " bonds," and on the bodily side with neurone arcs ; in the remainder of this chapter the word " bonds " will be used. But there is no necessity beyond that of convenience and vividness in this. The " bonds " spoken of may be identified by different readers with different entities. All a " bond " means, is some very simple aspect of the causal background. Some of them may be inherited, some may be due to education. There is no implication that the combined action of a number of them is the mere sum of their separate actions. There is no commitment to " mental atomism."

If, now, we have a causal background comprising innumerable bonds, and if any measurement we make can be influenced by any sample of that background, one measurement by this sample and another by that, all samples being possible ; and if we choose a number of different measurements and find their intercorrelations, the matrix of these intercorrelations will tend to be hierarchical, or at least tend to have a low reduced rank. This has nothing to do with the mind : it is simply a mathematical necessity, whatever the material used to illustrate it.

4. *A mind with only six bonds.*—We shall illustrate this fact first by imagining a " mind " which can form only six " bonds," which mind we submit to four " tests " which are of different degrees of richness, the one requiring the joint action of five bonds, the others of four, three, and two respectively (Thomson, 1927*b*). These four tests will (when we give them to a number of such minds) yield correlations with one another. For we shall suppose the

different minds not all to be able to form all six of the possible bonds, some individuals possessing all six, others possessing smaller numbers.

We have only specified the richness of each test, but have not said *which* bonds form each ability. There may, therefore, be different degrees of overlap between them, though some will be more frequent than others if we form all the possible sets of four tests which are of richness five, four, three, and two. If we call the bonds *a*, *b*, *c*, *d*, *e*, and *f*, then one possible pattern of overlap would be the following:

Test	*Bonds*					
1	*a*	*b*	*c*	*d*	*e*	.
2	.	*b*	*c*	*d*	*e*	.
3	.	.	.	*d*	*e*	*f*
4	.	.	*c*	*d*	.	.

If we for further simplicity suppose these bonds to be equally important, and use the formula—

$$\text{Correlation} = \frac{\text{overlap}}{\text{geometrical mean of the two totals}}$$

we can calculate the correlations which these four tests would give, namely:

	1	2	3	4
1	.	$\frac{4}{\sqrt{20}}$	$\frac{2}{\sqrt{15}}$	$\frac{2}{\sqrt{10}}$
2	$\frac{4}{\sqrt{20}}$	.	$\frac{2}{\sqrt{12}}$	$\frac{2}{\sqrt{8}}$
3	$\frac{2}{\sqrt{15}}$	$\frac{2}{\sqrt{12}}$	.	$\frac{1}{\sqrt{6}}$
4	$\frac{2}{\sqrt{10}}$	$\frac{2}{\sqrt{8}}$	$\frac{1}{\sqrt{6}}$	.

and we notice that all three tetrad-differences are zero. However, if we picked our four tests at random (taking

care only that they were of these degrees of richness) we would not always or often get the above pattern: in point of fact, we would get it only 12 times in 450. Nevertheless, it is one of the most probable patterns. In all, 78 different patterns of the bonds are possible—always adhering to our five, four, three, and two—the probability of each pattern ranging from 12 in 450 down to 1 in 450. One of the two least-probable patterns is the following:

Test	*Bonds*					
1	*a*	*b*	*c*	*d*	*e*	.
2	*a*	*b*	*c*	.	.	*f*
3	.	.	.	*d*	*e*	*f*
4	.	.	.	*d*	*e*	.

This pattern gives the correlations:

	1	2	3	4
1	.	$\frac{3}{\sqrt{20}}$	$\frac{2}{\sqrt{15}}$	$\frac{2}{\sqrt{10}}$
2	$\frac{3}{\sqrt{20}}$	.	$\frac{1}{\sqrt{12}}$	0 .
3	$\frac{2}{\sqrt{15}}$	$\frac{1}{\sqrt{12}}$	.	$\frac{2}{\sqrt{6}}$
4	$\frac{2}{\sqrt{10}}$	0	$\frac{2}{\sqrt{6}}$	.

This time the tetrads are not zero, but—

$$-\frac{2}{\sqrt{120}} \qquad \frac{4}{\sqrt{120}} \qquad \frac{6}{\sqrt{120}}$$

It is possible in this way to calculate the tetrad-differences for each one of the 78 possible patterns of overlap which can occur. When we then multiply each pattern by the expected frequency of its occurrence in 450 random choices of the four tests, we get 450 values for each tetrad-difference, distributed as follows:

Values of $F \times \sqrt{120}$	*Frequency of* F_1	F_2	F_3
8			2
7		4	0
6		8	14
5	9	2	6
4	27	34	23
3	6	12	30
2	75	72	48
1	61	66	72
0	99	54	81
− 1	56	78	36
− 2	67	42	42
− 3	16	30	60
− 4	30	36	18
− 5	0	0	0
− 6	4	12	18
	450	450	450

Although the distribution of each F about zero is slightly irregular, the average value of each F is *exactly* zero. For F_1 the variance is—

$$\sigma^2 = \frac{2{,}164}{120 \times 450} = \cdot 040$$

We see, then, that in this universe of very primitive-minded men, whose brains can form only six bonds, four tests which demanded respectively five, four, three, and two bonds would give tetrad-differences whose expected value would be zero, the values actually found being grouped around zero with a certain variance. There is no particular mystery about the four "richnesses" five, four, three, and two, by the way. We might have taken any four "richnesses" and got a similar result. If we performed the still more laborious calculation of taking all possible kinds of four tests, we should have obtained again a similar result. If there are no linkages among the bonds, the most probable value of a tetrad-difference will always

be zero; and if all possible combinations of the bonds are taken, the average of all the tetrad-differences will be zero. With only six bonds in the "mind," however, the scatter on both sides of zero will be considerable, as the above value of the standard deviation of F_1 shows, viz.—

$$\sigma = \sqrt{\cdot 040} = \cdot 20$$

5. *A mind with twelve bonds.*—But as the number of bonds in the mind increases, the tetrad-differences crowd closer and closer to zero. Let us, for example, suppose exactly the same experiment as above conducted in a universe of men whose minds could form twelve bonds (instead of six), the four tests requiring ten, eight, six, and four of these (instead of five, four, three, and two) (Thomson, 1927*b*). This increase in complexity enormously increases the work of calculating all the possible patterns of overlap, and the frequency of each. There are now 1,257 different square tables of correlation coefficients and still more patterns of overlap, some of which, however, give the same correlations. When each possibility is taken in its proper relative frequency (ranging from once to 11,520 times) there are no fewer than 1,078,110 instances required to represent the distribution. They have, nevertheless, all been calculated, and the distribution of F_1 was as follows:

$\sqrt{1920}\ F_1$	*Freq.*	$\sqrt{1920}\ F_1$	*Freq.*	$\sqrt{1920}\ F_1$	*Freq.*	$\sqrt{1920}\ F_1$	*Freq.*
20	225	7	17,760	− 3	31,432	− 13	624
18	1,800	6	74,392	− 4	72,676	− 14	8,792
16	1,755	5	15,744	− 5	53,808	− 15	4,144
15	4,600	4	52,085	− 6	49,828	− 16	8,970
14	8,840	3	121,608	− 7	21,240	− 18	112
12	19,610	2	42,384	− 8	41,951	− 19	456
11	10,682	1	28,096	− 9	5,896	− 20	584
10	8,860	0	122,699	− 10	29,184	− 24	28
9	26,696	− 1	63,024	− 11	8,960		
8	87,735	− 2	81,208	− 12	15,672		
						Total	1,078,110

This table again gives an average value of F_1 exactly equal to zero. But the separate values of the tetrad-

difference are grouped more closely round zero than before, with a variance now given by—

$$\sigma^2 = \frac{37,166,400}{1,920 \times 1,078,110} = 0{\cdot}018$$

This is rather less than half the previous variance. Doubling the number of bonds in the imagined mind has halved the variance of the tetrad-differences. If we were to increase the number of potential bonds supposed to exist in the mind to anything like what must be its true figure, we would clearly reach a point where the tetrad-differences would be grouped round zero very closely indeed.

The principle illustrated by the above concrete example can be examined by general algebraic means, and the above suggested conclusion fully confirmed (Mackie, 1928*a*, 1929). It is found that the variance of the tetrad-differences sinks in proportion to $1/(N - 1)$, where N is the number of bonds, when N becomes large, and the above example agrees with this even for such small N's as 6 and 12: for—

$$\frac{6 - 1}{12 - 1} \times {\cdot}040 = {\cdot}018 \quad \text{as found.}$$

In this mathematical treatment, bonds have been spoken of as though they were separate atoms of the mind, and, moreover, were all equally important. It is probably quite unnecessary to make the former assumption, which may or may not agree with the actual facts of the mind, or of the brain. Suitable mathematical treatment could probably be devised to examine the case where the causal background is, as it were, a continuum, different proportions of it forming tests of different degrees of richness. And as for the second assumption, it is in all likelihood merely formal. Let the continuum be divided into parts of equal importance, and then the number of these increased and their extent reduced, keeping their importance equal. What is necessary, to give the result that zero tetrads are so highly probable, is *that it be possible to take our tests with equal ease from any part of the causal background; that*

there be no linkages among the bonds which will disturb the random frequency of the various possible combinations; in other words, that there be no "faculties" in the mind. And it is also necessary that all possible tests be taken in their probable frequency.

In any actual experiment, of course, it is quite impracticable to take all possible tests, which are indeed infinite in number. A sample of tests is taken. If this sample is large and random, then there should, in a mind without separate "faculties," without linkages between its bonds, be an approach to zero tetrads. The fact that this tendency attracted Professor Spearman's attention, and was sufficiently strong to make him at first believe that all samples of tests showed it, provided care was taken to avoid tests so alike as to be almost duplicates (which would be "statistical impossibilities" in a random sample), indicates that the mind is indeed very free to use its bonds in any combination, that they are comparatively unlinked.

6. *Professor Spearman's objections to the sampling theory.*—A theory very similar to that of the sampling theory (but, as will be explained, with an entirely different meaning of sampling) had previously been considered by Professor Spearman (Spearman, 1914, 109 footnote), but had been dismissed by him because it would give a correlation between any two columns of the correlation matrix equal to the correlation between the two variates from which the columns derived, both of which correlations (he added) would on this theory average little more than zero (see also Spearman, 1928, Appendices I and II). A further objection raised by him (*Abilities*, 96) is that the "doctrine of chance," as he calls the sampling theory, would cause every individual to tend to equality with every other individual, than which, as he said, anything more opposed to the known facts could hardly be imagined.

These conclusions, however, have been deduced from a form of sampling, if it can be called sampling, which differs from that proposed by the present writer in the sampling theory. In the "doctrine of chance" discussed by Spearman, each ability is expressed by an equation containing every one of the elementary components or bonds, each

with a coefficient or loading (see Thomson, 1935*b*, 76; and Mackie, 1929, 30). The different abilities differ only in the loadings of the "bonds," and although some of these may be zero, the number of such zero loadings is insignificant.

But the sampling theory assumes that each ability is composed of *some but not all* of the bonds, and that abilities can differ very markedly in their "richness," some needing very many "bonds," some only few. It further requires some approach to "all-or-none" reaction in the "bonds"; that is, it supposes that a bond tends either not to come into the pattern at all, or to do so with its full force. This does not seem a very unnatural assumption to make. It would be fulfilled if a "bond" had a threshold below which it did not act, but above which it did act; and this property is said to characterize neurone arcs and patterns. When this form of sampling is assumed—and it is submitted that this is the normal meaning of sampling—then neither do the correlations become zero with an infinity of bonds, nor men equal; but the rank of the correlation matrix tends to be reducible to a small number, *if all possible correlations are taken*, and finally to be *one* as the bonds increase without limit.

It is important to realize what is meant by the rank *tending* to rank 1 as more and more of the possible correlations are taken. When the rank is 1 the tetrad-differences are zero. But clearly, the reader may say, taking more and more samples of the bonds to form more and more tests will not change in any way the pre-existing tetrad-differences, will not make them zero if they are not zero to start with. That is perfectly true; but that is not what is meant. As more and more tests are formed by samples of the bonds, the number of zero and very small tetrads will increase and swamp the large tetrads. The sampling theory does not say that all tetrads will be exactly zero, or the rank exactly 1. It says that the tetrads will be distributed about zero (not because each is taken both plus and minus, but when all are given their sign by the same rule) with a scatter which can be reduced without limit, in the sense that with more bonds the *pro-*

portion of large tetrads becomes smaller and smaller; always provided all possible samples are taken, i.e. that the family of correlation coefficients is complete.

With a finite number of tests this, of course, is not the case; but if the tests are a random sample of all possible tests, there will again be the approach to zero tetrads. The same will be true if the tests are sampling not the whole mind, but some portion of it, some sub-pool of our mind's abilities. If we stray from this pool and fish in other waters, we shall break the hierarchy; but if we sampled the *whole* pool of a mind, we should again find the tendency to hierarchical order. If the mind is *organized* into sub-pools (such as the verbal sub-pool, say), then we shall be liable to fish in two or three of them, and get a rank of 2 or 3 in our matrix, i.e. get two or three common factors, in the language of the other theory.

7. *Contrast with physical measurements.*—The tendency for tetrad-differences to be closely grouped around zero appears to be stronger in mental measurements than elsewhere; stronger, for example, than in physical measurements (*Abilities*, 142–3). In the comparisons which have been made, there has been some injustice done to the physical distributions; for diagrams have been published showing all the larger tetrads lumped together on to a small base so as to make the distribution look actually U-shaped. If, however, equal units are used throughout, the tetrad-differences are seen to be distributed here also in a bell-curve centred on zero (Thomson, 1927*a*),* though with a variance a good deal larger than is found in mental measurements (especially, of course, when the latter have been purified of all tests which give large tetrad-differ-

* In the paper quoted (Thomson, 1927*a*), the author mistakenly took each tetrad-difference with the sign obtained by beginning in every case with the north-west element. It is, however, Professor Spearman's practice to take every tetrad-difference twice, once positive and once negative. If this be done, a histogram like that on page 249 of the paper quoted becomes, of course, perfectly symmetrical. This change could be made throughout the paper without in any way affecting its main argument. The figure on page 249 (Thomson, 1927*a*) should be compared with that on page 143 of *The Abilities of Man*.

ences !). In spite of the difficulty of arriving, therefore, at a fair judgment with such evidence, it seems nevertheless likely that physical measurements do indeed show a weaker tendency to zero tetrads. For the tendency to zero tetrads, outlined above, due to the measurements sampling a complex of many bonds, will show itself only when the measurements in a battery are a fairly random sample of all the measurements which might be made.

Now, in physical measurements this is not the case. We do not measure a person's body just from anywhere to anywhere. We observe organs and measure them—leg, cranium, chest girth, etc. The variates are not a random sample of all conceivable variates. In other words, the physical body has an obvious structure which guides our measurements. The background of innumerable causes which produce just this particular body which is before us cannot act in all directions, but only in linked patterns. The tendency to zero tetrad-differences in the mind is due to the fact that the mind has, comparatively speaking, no organs. We can, and do, measure it almost from anywhere to anywhere. No test measures a leg or an arm of the mind; every test calls upon a group of the mind's bonds which intermingles in most complicated ways with the groups needed for other tests, without being a set pattern immutably linked into an organ. Of all the conceivable combinations of the bonds of the mind we can, without great difficulty, take a random sample, whereas in physical measurements we take only the sample forced on us by the organs of the body. Being free to measure the mind almost from anywhere to anywhere, we can get a set of measurements which show "hierarchical order" without overgreat trouble. We can do so because the mind is so comparatively structureless. Mental measurements tend to show hierarchical order, and to be susceptible of mathematical description in terms of one general factor and innumerable specifics, not because there are specific neural machines through which its energy must show itself, but just exactly because there are no fixed neural machines. The mind is capable of expressing itself in the most plastic and Protean way, especially before education, language, the subjects of

the school curriculum, the occupation, and the political beliefs of adult life have imposed a habitual structure on it. It is not without significance that the " factor " most widely recognized after Spearman's g is the verbal factor v, the mother-tongue being, as it were, the physical body of the mind, its acquired structure.

8. *Interpretation of* g *and the specifics on the sampling theory.*—We saw in Chapter III that the fraction expressing the square of the saturation of a test with g expresses in the sampling theory the fraction of the whole mind, or of the sub-pool of the mind, which that test forms. If the hierarchical battery is composed of extremely varied tests, which cover very different aspects of the mind's activity, this fraction may be taken as being of the whole mind—of the whole mind, that is, of an ideal man who can perform all of these tests perfectly, and all others which can extend their hierarchy. When we estimate a person's g, from such a battery, we are deducing a number which expresses how far that person is above or below average in the number of these bonds which *his* mind can form. This interpretation of g agrees well with an opinion arrived at, from quite another line of approach, by E. L. Thorndike, who on and near page 415 of his *Measurement of Intelligence* enunciates what has been called by others the Quantity Hypothesis of intelligence—that one mind is more intelligent than another simply because it possesses more interconnections out of which it can make patterns.

The difference in point of view between the sampling theory and the two-factor theory is that the latter looks upon g as being part of the test, while the former looks upon the test as being part of g. The two-factor theory is therefore compelled to postulate specific factors to account for the remainder of the variance of the test, and has to go on to offer some suggestion as to what specific factors *are*—perhaps neural engines. The sampling theory simply says that the test requires only such and such a fraction of the bonds of the whole mind—the same fraction which, on the two-factor theory, g forms of the variance of the test. For it, specific factors are mere figments, which do not arise unless, as can be done, the mathematical

equations which represent the tests are so manipulated that there appears to be only one link connecting them all. The sampling theory does not make this transformation of the equations (see Appendix, paragraph 6). Those who do so, if they adhere to the interpretation that g means all the bonds of the whole mind, have to suppose that the whole mind first takes part in each activity, but that in addition a specific factor is concerned; which specific factor, since they have already invoked the whole mind, must be for them a second action of part of the mind annulling its former assistance—which is absurd. The two-factor equations then do not allow us to consider g as being all the bonds of the mind. They are *mathematically* equivalent to the sampling equations, but not psychologically or neurologically. To the holder of the sampling theory, the factors of the other view are statistical entities only, g an average (or a total) of all a man's bonds, a specific factor the contrast between performance in any particular test and a person's general ability (Bartlett, 1937, 101–2). As a manner of speaking, the two-factor theory appears to the author to be much more likely to "catch on" with the man in the street, but much more likely to lead to the hypostatization of mere mathematical coefficients. The sampling theory lacks the good selling-points of the other, but is comparatively free from its dangers, and seems much more likely to come into line, in due time, with physiological knowledge of the action of the nervous system.

9. *Absolute variance of different tests.*—It will be noted, too, that on the sampling theory the different tests will naturally have different variances, the "richer" tests having a wider scatter. This seems only natural. It is customary, at any rate in theoretical discussions, to reduce all scores in different tests to standard measure, thereby equalizing their variance. This seems inevitable, for there is no means of comparing the scatter of marks in two different tests. But it does not follow that the scatter would be really the same if some means of comparison were available. When the same test is given to two different groups we have no hesitation in ascribing a wider variance to the one or the other group, and it seems con-

ceivable that a similar distinction might mentally be made between the scores made by one group in two different tests. The writer is completely in accord with M. S. Bartlett when he says (Bartlett, 1935, 205): "I think many people would agree . . . that the variation in mathematical ability displayed even in a selected group such as Cambridge Tripos candidates cannot be altogether put down to the method of marking adopted by the examiners." We may put these mathematics marks into standard measure, and we may put the marks scored by the same group in, say, a form-board test, also into standard measure. But that does not imply that at bottom the two variances are equal, if only we had some rigorous way of comparing them. Our common sense tells us plainly that they are not equal in the absolute sense, though for many purposes their difference is irrelevant. It seems to be no defect, then, but rather a good quality, of the sampling theory to involve different absolute variances.

10. *A distinction between* g *and other common factors.*—The writer is inclined, as the earlier sections of this chapter imply, to make a distinction in interpretation between the Spearman general factor *g* and the various other common factors, mostly if not all of less extent than *g*, which have been suggested. When properly measured by a wide and varied hierarchical battery, *g* appears to him to be an index of the span of the whole mind, other common factors to measure only sub-pools, linkages among bonds. The former measures the whole number of bonds; the latter indicate the degree of structure among them.

Some of this "structure" is no doubt innate; but more of it is probably due to environment and education and life. Its expression in terms of separate uncorrelated factors suggests what is almost certainly not the case, that the "sub-pools" are separate from one another. The actual organization is likely to be much more complicated than that, and its categories to be interlaced and interwoven, like the relationships of men in a community, plumbers and Methodists, blonds, bachelors, smokers, conservatives, illiterates, native-born, criminals, and school-teachers, an organization into classes which cut

across one another right and left. No doubt these too could be replaced, and for some purposes replaced with advantage, by a smaller number of uncorrelated common factors and a large number of factors specific to plumbers, smokers, and the rest. But the factors would be pure figments. What the factorist calls the verbal factor, for example, is something very different from what the world recognizes as verbal ability. The latter is a compound, at least of g and v, and possibly of other factors. The v of the factorist is something uncorrelated with g, something which the person of low g is just as likely to have as the person with high g. Oblique factors are, it is true, envisaged by Thurstone, but, as has been said, probably only within sampling limits; that is, they are slightly distorted orthogonal factors.

Further, it is improbable that the organization of each mind is the same. The phrase "factors of *the* mind" suggests too strongly that this is so, and that minds differ only in the amount of each factor they possess. It is more than likely that different minds perform any task or test by different means, and indeed that the same mind does so at different times.

Yet with all the dangers and imperfections which attend it, it is probable that the factor theory will go on, and will serve to advance the science of psychology. For one thing, it is far too interesting to cease to have students and adherents. There is a strong natural desire in mankind to imagine or create, and to name, forces and powers behind the façade of what is observed, nor can any exception be taken to this if the hypotheses which emerge explain the phenomena as far as they go, and are a guide to further inquiry. That the factor theory has been a guide and a spur to many investigators cannot be denied, and it is probably here that it finds its chief justification.

CHAPTER XIX

"STOP-PRESS"

1. *Recent publications, and three questions.*—Since it is inevitable that, after the manuscript of a scientific book has been sent to the publishers, articles and books should appear to which it is desirable to refer, it was arranged that this postscript should be written as late as possible during the printing, to enable the "latest news" to be incorporated without interference with the body of the book. The two most interesting and relevant of the publications which have appeared during the printing are probably Thurstone's monograph *Primary Mental Abilities* (Thurstone, 1938) and Burt's paper *The Analysis of Temperament* (Burt, 1938). A comparison of these two as regards their underlying principles will be a convenient way of discussing in a few final paragraphs what appear to be the main theoretical questions needing an answer, namely:

(1) What metric or system of units is to be used in factorial analysis?

(2) On what principle are we to decide where to stop the rotation of our factor-axes or how to choose them so that rotation is unnecessary?

(3) Is the principle of minimizing the number of common factors, i.e. of analysing only the "communal" variance, to be retained?

Thurstone wholeheartedly accepts, indeed has been mainly responsible for, the third principle. With regard to metric, he calls the variance of each test unity, and analyses *correlations*, not covariances; but it will be shown below that the question of what metric to use is really unimportant to anyone accepting the principle of the fewest common factors, an argument in favour of the latter. He holds very strongly that unrotated factors, as for example those first obtained by the use of the "centroid" process, are not psychologically significant, and he rotates

all the centroid common factors (but not the specifics) until what he calls "simple structure" is approximated to. His method is dominated by the two concepts of "fewest common factors" and "simple structure."

Burt's method is dominated by a totally different idea, namely the desire that the factors arrived at by analysing persons should be the same as those arrived at by analysing traits, or more exactly that the factors and loadings of the one kind of analysis should be the loadings and factors of the other. He can only attain this if he somehow obtains a matrix of marks which is centred both ways, that is, whose columns and rows both add to zero; and if he analyses actual variances and covariances, not correlations. This involves adopting a different set of units, a different metric, from that of unit standard deviations, but the units he actually adopts are, it would seem, arbitrary and fortuitous, and the present author will later in this chapter suggest a "natural metric" based upon the sampling theory.

As to rotations, Burt makes none. His plan may be crudely described, with reservations, as removing and then disregarding the first centroid factor (with full variances, not communalities), and then using the larger principal components of the residues as his significant factors. There are several statistical difficulties, which he is probably wise to pass over lightly in order to present the main idea as vividly as possible.

The fact that on these apparently incompatible principles each author arrives at factors which seem to him to possess psychological meaning is either rather disquieting, as suggesting that each is unconsciously allowing himself to find what he would like to find, or rather encouraging, as suggesting that in spite of their differences the two methods are working towards a common end. Comparison is made difficult, and somewhat unfair to the two authors, by the fact that Burt was analysing temperaments, and Thurstone abilities.

2. *Primary mental abilities and* "*g.*"—Thurstone administered 57 tests, requiring 15 hours, to some 240 students of Chicago who volunteered. The centroid analysis with

guessed communalities gave twelve factors, which were then rotated into an approximation to simple structure.

In performing the rotations Thurstone, after using more complicated methods (not without success), nevertheless reverted to the simple plan of rotating two factors at a time graphically. This is done by plotting the loadings of two factors as co-ordinates on squared paper, and rotating the axes by inspection so as to reduce the number of negative loadings and produce many zeros (compare Figure 17 and Section 10 of Chapter IV). The new loadings could be found from the diagram by measurement, but more accurately are calculated by postmultiplying the two columns of loadings by the orthogonal matrix:

$$\begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}$$

Thus if the diagram had Factor I as the vertical and Factor II as the horizontal axis, and the rotation decided upon were one of $-30°$, i.e. $30°$ in a clockwise direction, then the loadings for the new factors would be—

$$\begin{array}{ll} \text{for } \mathrm{I}_1, & A\cos(-30°) - B\sin(-30°) \\ \text{for } \mathrm{II}_1, & A\sin(-30°) + B\cos(-30°) \end{array}$$

where A and B are the loadings of the original Factors I and II. Each of the rotated factors may afterwards be paired with other factors and again rotated. The process converges, it appears, to the same result as that obtained in more complicated mathematical ways.*

Of the twelve factors after rotation,† Thurstone feels considerable confidence in naming the first seven as S spatial, P perceptual, N numerical, V verbal relations,

* It is not quite clear from the monograph whether the actual criterion used in the rotations was the abolition of negative loadings or the maximizing of the number of zero loadings. Page 71 suggests that both were taken into consideration, whereas page 72 says that the zeros were maximized and the negatives then spontaneously disappeared. Possibly these statements refer to different methods, for the rotations were done more than once.

† A thirteenth factor appears in the rotated table, but it has no loadings of any significant size and appears to be outside the space of the twelve original unrotated factors.

M memory, W words (i.e. single words), and I induction. Two others he tentatively names R reasoning and D deduction. For the other three he can find no clear psychological meaning. None of these factors is a general factor. The most extensive is the factor V, which has 15 loadings, out of the 57, which are definitely non-zero. On this question of a general factor Thurstone writes: "As far as we can determine at present, the tests that have been supposed to be saturated with the general common factor divide their variance among primary factors that are not present in all the tests. We cannot report any general common factor in the battery of tests that have been analysed in the present study." He had included Spearman's *Figure Classification* test in the battery as one of the best tests for *g*. Its analysis* in his final table is—

$$\cdot 393S + \cdot 405I + \cdot 398D$$
$$+ \text{ factors with smaller loadings } + \cdot 585 \text{ specific}$$

In view of Thurstone's rules for finding simple structure, which involve maximizing the number of zero loadings, it would indeed appear very unlikely for a general factor to remain. On his page vii, however, Thurstone remarks: "Our methods do not preclude it. The presence of a general factor could be indicated by a large part of the communality of each test that remains unaccounted for by the common factors that can be identified in a simple structure." This appears to mean that when, as in the present experiment, there are three unnamed factors, and two others whose significance is doubtful, these three (or five) might then be rotated in their own space away from simple structure till a general factor reappears. Whatever the exact details of Thurstone's meaning, it is clear, and it is very interesting to notice, that a general factor will crop up in his system, if at all, only *after* the identification of the psychologically significant common, but not general, fac-

* The spatial factor S is not surprising, since the test is composed of geometrical figures. What, however, *is* surprising is that Test 6, Verbal Classification, introduced to parallel Spearman's Figure Classification but with verbal material, has a slightly *higher* saturation with this spatial factor.

tors in a simple structure.* This is the very reverse of the practice of the Spearman school, of removing the general factor *g* first; and the reverse too of Burt's method, now to be described, in which a general factor (though not *g*) is removed at the outset.

8. *An analysis of temperaments.*—Burt's paper is concerned with applying the techniques of factorial analysis to emotional characteristics. The place of " tests " was here taken by assessments made by observers, on eleven traits such as anger, joy, sex, disgust, etc. Persevering in his desire to obtain factors and loadings which are identical in the analysis of persons and traits, Burt again starts from a matrix of marks which is centred both by rows and by columns, and analyses *covariances*. He does not, however, simply and crudely take raw marks and centre them both ways. What he actually does is connected with the idea of removing the average, considered as a general factor, and analysing the residual covariances. There are serious mathematical questions raised by his procedure.

As we have said, to attain his aim Burt must have a set of marks centred both ways. Now, the matrix of covariances calculated from such a doubly centred set of marks is itself double-centred (see those at the end of Chapter XIV, Section 1, page 214). But such a doubly centred matrix of covariances also occurs in the residues of the "centroid" process (see Chapter II, Section 5, page 28), which is indeed why the device of temporary sign changing had to be there adopted. If, therefore, we could somehow do something legitimate and understandable to the original

* At the very last moment of proof-reading, an analysis of the same data by Holzinger and Harman on the Bifactor method comes to hand (1938, *Psychometrika*, *3*, 45–60). They find an important general facter due, as they truly say, " to our hypothesis of its existence and the essentially positive correlations throughout." With this important difference, their analysis shows several resemblances to that of Thurstone. In the same number, Thurstone analyses some of the same tests on different subjects and claims that the same factors emerge. There is no doubt in the present writer's mind that some agreement is present in these three analyses, though it seems to him unwise to exaggerate it. Numerically the loadings of ten differ by large amounts.

raw marks to modify them so that they would give the first centroid residues direct, both for traits and persons, we would have attained the end which Burt seeks.

Now, the matrix of first residues in the "centroid" process, if we use correlations, and unity in each diagonal cell, is the matrix of partial correlations for constant average standard score. If we use covariances and full variances, it is the matrix of partial covariances for constant average raw score. Speaking first of covariances between *traits*, a calculated partial covariance for constant average score is the same thing (if the distributions are normal) as the actual covariance found in a subpopulation of persons each of whom has in fact the same average score in all the traits. Burt therefore selects such a subpopulation. From 500 cases he chose 124 whose average mark for the emotional traits was, *for each person*, approximately the same as the average of the entire 500. Since these 124 are all alike in average, measuring each person's score from his own average in all the traits will not change the covariances between traits. So far Burt is on perfectly sound ground.

But from this matrix he also needs to calculate the covariances between persons ; and unless the traits also turn out to have each the same average over the sub-group of persons, centring the marks by traits will distort the covariances between persons and make them meaningless. Now, in general one would not expect the subpopulation of persons, equal in average trait score, to give a matrix of marks in which the traits also had equal averages over the persons. In the whole population one could, of course, ensure such equality of trait averages over the persons by instructing the observers to distribute their marks symmetrically in each trait about the same conventional average, and this Burt did. But in the subpopulation these trait averages would not remain equal unless each trait were equally correlated with the average of all, which is in itself unlikely, and seems in the present instance, as far as calculations on the data given would lead one to judge, to have been very far from being the case. Yet Burt tells us that the trait averages diverged only slightly from one another ; this statement, however, is not made about the

subpopulation of 124, but about eleven children chosen from them to be his actual group to be analysed. It would have been an advantage to have had more particulars about the 500 and the 124.

These eleven children were picked so that their trait correlations were practically identical with those of the larger group (presumably the 124). The number 11 was arrived at through the fact that the 124 cases appeared to fall into 13 types, of which two rare types were rejected, which disturbed the correlations. It is to be noted that as there were also eleven traits, the final matrix of marks is square, which has certain mathematical consequences. The whole analysis is, one would think, a very special case. And since it is directed throughout by the desire to have an analysis which would be also arrived at by interchanging persons and traits, one would like to consider more carefully than is here possible, or to find by trial, whether, starting with eleven children and 500 traits, the selection of 124 or so traits, and then eleven, on a parallel plan to the above, would have left us with an 11×11 matrix just like that actually reached. However, these questions are only the criticisms of an admirer, and the valiant effort to put into actual practice a theoretical principle (of reciprocity between person factors and trait factors) is most noteworthy.

4. *The average as a general factor.*—Leaving aside these particular doubts and criticisms, we notice that first of all Burt extracts a general factor which is the average performance. This will, of course, vary with the battery of tests or traits, in answer to which objection it may be said that if the battery be large and varied, the average will not relatively alter much with additions or changes. This general average is not Spearman's *g*. It "takes out" more variance than that, and involves negative loadings in the later factors which cannot be removed by rotation of those later factors alone. The presence of negative loadings in Burt's present experiment does not occasion disquietude, because in emotional traits there is not that opposition to their presence that many psychologists feel in the case of factors on the intellectual side of the mind.

The first centroid factor of Thurstone's process is also an average, though since communalities are used it is not an average of the whole scores. But Thurstone at once rotates all the common factors, including this first average, into a new position. Burt rotates neither the average nor his later factors, which are the principal components of the doubly centred matrix (which has one dimension missing), but accepts them as they stand. It would almost seem correct to describe Burt's aim as the more modest one of merely describing the actual marks—he himself uses phrases which seem to imply this—and not the more ambitious one of reaching factors which have a kind of independent existence and will be invariant in different batteries.

5. *The use of covariances.*—Burt's chief reason for using covariances instead of correlations is no doubt that only then can a simple relation between trait factors and person factors be stated.* But this use of variances and covariances commits him to a metric. His method of analysis, after he has obtained his doubly centred matrix, is equivalent to finding the principal axes of the ellipsoid of density and using them as factors. Now, this ellipsoid must exist in some space or other. If we analyse correlations we are using a space in which the standard deviations of all variables are alike, admittedly a confession of ignorance. There is undoubtedly something to be said for the probability of real differences of standard deviation existing (see Chapter XVIII, Section 9). In that case, if we knew these real standard deviations, we would use variances and covariances and the space corresponding to them (compare Hotelling, 1933, 421–2 and 509–10). But it surely cannot be right to use a space whose metric is dependent upon accidental and irrelevant differences of variance in the variables. In Burt's experiment, for example, the traits

* He gives also other reasons, one of which is surely erroneous. For he holds that dividing the covariances by standard deviations to obtain correlations gives an unwarranted weight to correlations concerning any trait with an *artificially* small scatter. But in that case the covariance is already too small, for the same reason. It is the correlation, not the covariance, which is independent of the standard deviations.

sex and anger have (unaveraged) variances of 588 and 6,818 respectively. He himself urges that this difference is due, not to a real difference in these traits, but to the teachers' ignorance of the sexual propensities of the children, as a result of which they " mark nearly every child near the average. On the other hand, bad and good temper . . . can scarcely be missed: marks (for anger) . . . therefore . . . exhibit an extremely wide range." The important point in this is that the differences in variances are not real differences. Yet in Burt's form of analysis the factors and their loadings depend on these accidental variances. It is true that Burt gives, in addition to the analysis of covariances, *an* analysis of the correlations. But it is not the principal components analysis of the correlations but a conventional reduction of the principal components of the covariances, so that even his analysis of the correlations gives factors which depend on the accidental variances.

6. *Natural variances and units.*—It would seem necessary, if variances and covariances are to be analysed, to have some system of natural units. Hotelling has already suggested one such, based upon the idea of the principal components of all possible tests; but it would seem to be an unattainable ideal (Hotelling, 1933, 510). A somewhat similar but not, it would appear, identical method, which is in some measure and in some situations actually attainable, can be based on the ideas of the sampling theory and has already been foreshadowed in Chapter XVIII, Section 9. Tests quite naturally have different variances on that theory, since they comprise larger or smaller samples of the " bonds " of the mind (see Thomson, 1935*b*, 87). In a hierarchical battery these natural variances are measured by the " coefficient of richness " (Chapter III, Section 2, page 45). The " richness " of Test k is given by—

$$\frac{r_{ik}r_{jk}}{r_{ij}},$$

the same quantity as the square of Spearman's " saturation with *g*." It is also, on the sampling theory, the fraction which the test forms of the pool of bonds which is being sampled, and is the natural variance of the test. In other

words, in a hierarchical battery the "saturation with g" of Spearman's theory is the "natural standard deviation" of the sampling theory. In a nearly hierarchical battery it can be estimated by Spearman's formula (Chapter IX, Section 5, page 154)—

$$\sqrt{\frac{A^2 - A'}{T - 2A}}$$

In a battery which is definitely not hierarchical, the same formula will nevertheless give a rough estimate of the natural standard deviation of each test. The general principle is that tests which show the most total correlation have the largest natural variance.

7. *Simple structure and units.*—Spearman's analysis of a hierarchical battery, and, in general, analyses made on the principle of the fewest common factors and rotation to simple structure, as are Thurstone's, are, however, independent of the standard deviations employed. Exactly the same result is arrived at whether correlations or covariances are used, except, of course, that the "saturations" of the correlational simple structure have to be multiplied in each test by the standard deviation of that test to give the "loadings" of the covariance simple structure.* The first centroid analysis depends on the variances used, but the differences disappear when simple structure is reached.

This independence of the units used does not hold for every kind of analysis. The loadings of a principal-axes analysis of covariances, when divided by the standard deviations, do not become the saturations of a principal-axes analysis of the corresponding correlations, though they can be rotated into these. It is a strong argument in favour of simple structure.

8. *Indeterminacy of minimal-rank analyses.*—On the other hand, all factors obtained in minimal-rank analyses, based on the principle of the fewest common factors, are to a greater or less extent indeterminate, because they outnumber the tests. Burt's system has here the advantage,

* This convenient distinction between *saturation* and *loading* is proposed by Burt in his last article, and would have been used in the body of this book had that article appeared in time.

since he has no specifics, and his factors need not be estimated but can be exactly calculated—where "exactly" means with the same exactness as that of the data—whereas estimations of Thurstone's factors are less exact than the data. The determinate and the indeterminate parts of each of Thurstone's factors in *Primary Mental Abilities* can be calculated by postmultiplying Table 7 on his page 98 by Table 8 on his page 96. We find :

Factor	*Variance of the Estimated Part*	*Variance of the Indeterminate Part*
S . . .	·611	·389
P . . .	·616	·384
N . . .	·825	·175
V . . .	·662	·338
M . . .	·431	·569
W . . .	·439	·561
I . . .	·397	·603
R . . .	·600	·400
D . . .	·519	·481

In three cases less than half of the factor variance has been estimated. The average for the nine factors is 56½ per cent. of the variance estimated. In other words, the factor estimates have large probable errors, in some cases as large as the estimates themselves. This has serious consequences for the utility of the whole system, which are not to be overcome by more reliable tests. The weakness is due to the excess of factors over tests, and this in turn is due to the principle of minimizing the number of common factors. It could only be overcome by discovering a battery of tests whose correlation matrix with unit diagonal cells was already of low rank, unless one were content to have as many common factors as tests (and give up the concept of few common factors), though one might then use only the larger ones. The conflict is between the ideas (*a*) of reproducing the correlations accurately with few factors, and (*b*) of reproducing the whole test variance accurately with few factors.

MATHEMATICAL APPENDIX

MATHEMATICAL APPENDIX

PARAGRAPHS

1. *Textbooks on matrix algebra.*—Some knowledge of matrix algebra is assumed, such as can be gained from the mathematical introduction to L. L. Thurstone's *The Vectors of Mind* (Chicago, 1935); Turnbull and Aitken's *Theory of Canonical Matrices*, Chapter I (London and Glasgow, 1932); H. W. Turnbull's *The Theory of Determinants, Matrices, and Invariants*, Chapters I–V (London and Glasgow, 1929); and M. Bôcher's *Introduction to Higher Algebra*, Chapters II, V, and VI (New York, 1936).

2. *Matrix notation.*—Let X be the matrix of raw scores of p persons in n tests, with n rows and p columns; and

when normalized by rows, let it be denoted by Z. The letters z and Z in the *text* of this book mean *standardized* scores, which are used in practical work, but in this appendix they mean *normalized* scores, so that—

$$ZZ' = R \quad . \quad . \quad . \qquad (1)$$

the matrix of correlations between n tests.

For many purposes it is convenient to think of solid matrices like Z as column (or row) vectors of which each element represents a row (or column). Thus Z can be thought of as a column vector z, of which each element represents in a collapsed form a row of a person's scores. Thus with three tests and four persons—

$$z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & z_{13} & z_{14} \\ z_{21} & z_{22} & z_{23} & z_{24} \\ z_{31} & z_{32} & z_{33} & z_{34} \end{bmatrix} = Z. \qquad (2)$$

In the theory of mental factors each score is represented as a loaded sum of the normalized factors f, the loadings being different for each test, i.e.—

$$z = Mf \text{ (specification equations)} \quad . \qquad (3)$$

where M is the matrix of loadings, and f the vector of v factors, collapsed into a column from F, the full matrix, of dimensions $v \times p$.

We note that p = number of persons,
n = number of tests,
v = number of factors.

The dimensions of M are $n \times v$. Equation (3) represents n simultaneous equations, and the form $Z = MF$ represents np simultaneous equations.

We now have—

$$R = ZZ' = (MF)(MF)' = MFF'M \quad . \qquad (4)$$

If the factors are orthogonal, we have—

$$FF' = I \quad . \quad . \quad . \quad . \quad . \qquad (5)$$

the unit matrix, and therefore—

$$R = MM' \quad . \quad . \quad . \quad . \quad . \qquad (6)$$

The resemblance in shape between this and—

$$R = ZZ' \quad . \quad . \quad . \quad . \quad . \qquad (1)$$

leads to a parallelism between formulæ concerning persons and factors (Thomson, 1935*b*, 75; Mackie, 1927, 74, and 1929, 34).

3. *Spearman's Theory of Two Factors* assumes that M is of the special form—

$$M = \begin{bmatrix} l_1 & m_1 & . & . & . \\ l_2 & . & m_2 & . & . \\ . & . & . & . & . \\ l_n & . & . & . & m_n \end{bmatrix}, \; l^2 + m^2 = 1 \qquad (7)$$

and therefore—

$$R = ll' + M_1^2 \quad . \quad . \quad . \quad . \quad (8)$$

where M_1 is the diagonal matrix which forms the right-hand end of M, and l is the first column of M. In this form it is clear that R is of rank 1 except for its principal diagonal. Its component ll' is the "reduced correlational matrix" of the Spearman case, and is entirely of rank 1. The elements l_1^2, l_2^2, . . . l_n^2, which form the principal diagonal of ll', are called "communalities."

4. *Multiple common factors.*—When more than one common factor is present, M takes the form—

$$M = [M_0 \vdots M_1] \quad . \quad . \quad . \quad . \quad (9)$$

where M_0 is the matrix of loadings of the common factors, represented in the Spearman case by the simple column l. We have then—

$$R = MM' = M_0M_0' + M_1^2 \quad . \quad . \quad . \quad (10)$$

where the "reduced correlation matrix" M_0M_0' is of rank r, the number of *common* factors, and is identical with R except for having "communalities" in its principal diagonal.

5. *Orthogonal rotations.*—If we express the v factors f in terms of w new factors φ by the equation—

$$f = A\varphi \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where A is a matrix of v rows and w columns, we have—

$$z = Mf = MA\varphi \quad . \quad . \quad . \quad (12)$$

an expression of the tests z as linear loaded sums of a different set of factors, with a matrix of loadings MA.

If—

$$AA' = I \quad . \quad . \quad . \quad . \quad . \quad (13)$$

the new factors φ are orthogonal like the old ones. They can be as numerous as we like, but not less than the number of tests unless the matrix R is singular. (12) represents a rigid rotation of the orthogonal axes f into new positions, with dimensions added or abolished.

6. *The sampling theory.*—The following transformation is of interest as showing the connexion between the Theory of Two Factors and the Sampling Theory (Thomson, 1935*b*, 85). We shall write it out for three tests only, but it is quite general. Consider the orthogonal matrix:

$$\begin{bmatrix} lll & mll & lml & llm & mml & mlm & lmm & mmm \\ mll & -lll & mml & mlm & -lml & -llm & mmm & -lmm \\ lml & mml & -lll & lmm & -mll & mmm & -llm & -mlm \\ llm & mlm & lmm & -lll & mmm & -mll & -lml & -mml \\ mml & -lml & -mll & mmm & lll & -lmm & -mlm & llm \\ mlm & -llm & mmm & -mll & -lmm & lll & -mml & lml \\ lmm & mmm & -llm & -lml & -mlm & -mml & lll & mll \\ mmm & -lmm & -mlm & -mml & llm & lml & mll & -lll \end{bmatrix} \quad (14)$$

wherein the omitted subscripts 1, 2, and 3 are to be understood as existing always in that order, so that *mll* means $m_1l_2l_3$.

If we take for A in Equation (12) the first four rows of this orthogonal matrix, and for M the Spearman form (7) with three tests, the result is to transfer to eight new factors, yielding:

$$\begin{aligned} z_1 &= l_2l_3\varphi_1 + m_2l_3\varphi_3 + l_2m_3\varphi_4 + m_2m_3\varphi_7 \\ z_2 &= l_1l_3\varphi_1 + m_1l_3\varphi_2 + l_1m_3\varphi_4 + m_1m_3\varphi_6 \quad . \quad . \quad (15) \\ z_3 &= l_1l_2\varphi_1 + m_1l_2\varphi_2 + l_1m_2\varphi_3 + m_1m_2\varphi_5 \end{aligned}$$

Each z is here in normalized units. If, however, we change to new units by multiplying the three equations by l_1, l_2, and l_3 respectively, we have:

$$\begin{aligned} l_1z_1 &= l_1l_2l_3\varphi_1 + l_1m_2l_3\varphi_3 + l_1l_2m_3\varphi_4 + l_1m_2m_3\varphi_7 \\ l_2z_2 &= l_1l_2l_3\varphi_1 + m_1l_2l_3\varphi_2 + l_1l_2m_3\varphi_4 + m_1l_2m_3\varphi_6 \quad . \quad . \quad (16) \\ l_3z_3 &= l_1l_2l_3\varphi_1 + m_1l_2l_3\varphi_2 + l_1m_2l_3\varphi_3 + m_1m_2l_3\varphi_5 \end{aligned}$$

and the variates l_1z_1, l_2z_2, and l_3z_3 are now susceptible of the explanation that each is composed of l_i^2N small equal

components drawn at random from a pool of N such components, all-or-none in nature. In that case $l_1^2 l_2^2 l_3^2 N$ components would probably appear in all three drawings (φ_1); $l_1^2 l_2^2 m_3^2 N$ components would probably appear in the first two drawings, but not in the third (φ_4); and so on down to $m_1^2 m_2^2 m_3^2$ components, which would not appear at all (φ_8, which is missing from the equations).

The transformation can, of course, be reversed, and the sampling theory equations converted into the two-factor equations.

7. *Hotelling's "principal components"* are the principal axes of the ellipsoids of equal density—

$$z'R^{-1}z = \text{constant} \quad . \quad . \quad . \quad . \quad (17)$$

when the *test* vectors are orthogonal axes (Hotelling, 1933). To find the principal axes involves finding the latent roots of R^{-1}. The Hotelling process consists of (*a*) a rotation of the axes from the orthogonal test axes to the directions of the principal axes; and (*b*) a set of strains and stresses along these new axes to standardize the factors, making the ellipsoid spherical and the original axes oblique. The transformation from the tests to the Hotelling factors γ being from Equation (3)—

$$z = M\gamma \quad (M \text{ square})$$

the ellipsoids (17) become—

$$\text{constant} = z'R^{-1}z = \gamma'(M'R^{-1}M)\gamma = \gamma'\gamma \quad . \quad (18)$$

since they become spheres. Therefore we must have—

$$M'R^{-1}M = I \quad . \quad . \quad . \quad (19)$$

The locus of the mid points of chords of $z'R^{-1}z$ whose direction cosines are h' is the plane $h'R^{-1}z = 0$, and if this is a principal plane it is at right angles to the chords it bisects, i.e.—

$$h'R^{-1} = \lambda h'$$

which has non-trivial solutions only for—

$$| R^{-1} - \lambda I | = 0$$

the roots λ of which are the "latent roots" of R^{-1}, while each h' is a "latent vector."

Now, if H is the matrix of normalized latent vectors of R^{-1}, we have—

$$H'R^{-1}H = \Lambda$$

where Λ is the diagonal matrix of the latent roots of R^{-1}; so that a solution for M corresponding to rotation to the principal axes and subsequent change of units to give a sphere is seen to be—

$$M = H\Lambda^{-\frac{1}{2}} \quad . \quad . \quad (20)$$

The latent vectors of R are the same as those of R^{-1}, or of any power of R, and Hotelling's process described in the text (Chapter V) finds the latent roots (forming the diagonal matrix D) and the latent vectors (forming H) of R. We then have—

$$M = HD^{\frac{1}{2}} \quad . \quad . \quad (21)$$

For the convergence of the process, see Hotelling's paper of 1933, pages 14 and 15.

Since in Hotelling analyses M is square, we can write—

$$\begin{aligned} \gamma &= M^{-1}z = (HD^{\frac{1}{2}})^{-1}z \\ &= D^{-\frac{1}{2}}H^{-1}z = D^{-1}(D^{\frac{1}{2}}H')z = D^{-1}M'z \quad . \quad (22) \end{aligned}$$

Each factor γ, that is, can be found from a *column* of the matrix M, divided by the corresponding latent root, used as loadings of the test scores z.

8. *The pooling square.*—If the matrix of correlations of $a + b$ variates is:

$$\left.\begin{array}{c|c} R_{aa} & R_{ab} \\ \hline R_{ba} & R_{bb} \end{array}\right. \quad . \quad . \quad . \quad (23)$$

and if the standardized variates a are multiplied by weights u, the standardized variates b by weights w, and each set of scores summed to make two composite scores, the resulting variances and covariances are:

$$\left.\begin{array}{c|c} u'R_{aa}u & u'R_{ab}w \\ \hline w'R_{ba}u & w'R_{bb}w \end{array}\right. \quad . \quad . \quad (24)$$

as can be seen by writing out the latter expressions at length. The battery intercorrelation is therefore—

$$\frac{u'R_{ab}w \text{ or } w'R_{ba}u}{\sqrt{(u'R_{aa}u \times w'R_{bb}w)}} \quad . \quad . \quad (25)$$

If weights are applied to *raw* scores, each applied weight must be multiplied by each pre-existing standard deviation, in (25).

If there is only one variate in the a team, (25) becomes—

$$\frac{w'r_{ba}}{\sqrt{(w'R_{bb}w)}} \quad . \quad . \quad . \quad (26)$$

where r_{ba} represents a whole column of correlation coefficients. The values of w for which this reaches its maximum value will satisfy the equation—

$$\frac{\delta}{\delta w} \frac{w'r_{ba}}{\sqrt{(w'R_{bb}w)}} = 0 \quad . \quad . \quad (27)$$

that is—

$$w = \text{a scalar} \times R_{bb}^{-1}r_{ba} \quad . \quad . \quad (28)$$

consistent with the ordinary method of deducing regression coefficients.

9. *The regression equation.*—If z_0 is the one variate in the a team, and z are the b team, and if—

$$\hat{z}_0 = w'z \quad . \quad . \quad . \quad . \quad (29)$$

we wish to make $S(z_0 - \hat{z}_0)^2$ a minimum, that is—

$$\frac{\delta}{\delta w} S(z_0 - w'z)^2 = 0$$

$$Sz_0z' = w'Szz'$$

$$w' = r_{ab}'R_{bb}^{-1}$$

$$\hat{z}_0 = r_{ab}'R_{bb}^{-1}z \quad . \quad . \quad . \quad (30)$$

If $\mathbf{R}$ is the matrix of correlations of all the tests including z_0, the regression estimate of any one of the tests from a weighted sum of the others is given by—

$$\text{determinant } \mathbf{R}_z = 0 \quad . \quad . \quad . \quad . \quad . \quad (31)$$

where $\mathbf{R}_z$ is $\mathbf{R}$ with the row corresponding to the variate to be estimated replaced by the row of variates.

10. *Regression estimates of factors.*—When in the specifications—

$$z = Mf \quad . \quad . \quad . \quad . \quad (8)$$

the factors outnumber the tests, they cannot be measured but only estimated. To all men with the same set of scores z will be attributed the same set of estimated factors $\hat{f}$, though their " true " factors may be different. The

20

regression method of estimation minimizes the squares of the discrepancies between $\hat{f}$ and f, summed over the men. The regression equation (31) will be for one factor f_i—

$$\begin{vmatrix} \hat{f}_i & z' \\ m_i & R \end{vmatrix} = 0 \quad . \quad . \quad . \quad . \quad . \quad (32)$$

where m_i is a column of M. Expanding, we have—

$$\hat{f}_i = m_i'R^{-1}z$$

and in general—

$$\hat{f} = M'R^{-1}z \quad . \quad . \quad . \quad . \quad (33)$$

or, separating the common factors and the specifics—

$$\hat{f}_0 = M_0'R^{-1}z. \quad . \quad . \quad . \quad (34)$$

$$\hat{f}_1 = M_1R^{-1}z \quad . \quad . \quad . \quad . \quad (35)$$

the latter of which shows that we know the *proportionate* weights for each specific (the rows of R^{-1}) even before we know whether that specific exists (Wilson, 1934*b*, 194). The matrix of covariances of the estimated factors is—

$$K = M'R^{-1}M = \begin{bmatrix} M_0'R^{-1}M_0 & M_0'R^{-1}M_1 \\ M_1R^{-1}M_0 & M_1R^{-1}M_1 \end{bmatrix} \quad . \quad . \quad (36)$$

a square idempotent matrix of order equal to the number of factors, but trace only equal to the number of tests.

For one common factor, (34) reduces to Spearman's estimate—

$$\hat{g} = \frac{1}{1+S}\Sigma\frac{r_{ig}z_i}{1-r_{ig}^2} \quad . \quad . \quad (34a)$$

where $$S = \Sigma\frac{r_{ig}^2}{1-r_{ig}^2}$$

while $K = M_0'R^{-1}M_0$ in (36) reduces to $S/(1+S)$, the variance of $\hat{g}$.

11. *Direct and indirect vocational advice.*—If z_0 is an occupation and z a battery of tests, the estimate of a candidate's occupational ability is—

$$\hat{z}_0 = r_0'R^{-1}z \quad . \quad . \quad . \quad . \quad (37)$$

where the r_0 are the correlations of the occupation with the tests. If z_0 can be specified in terms of the common factors of z, and a specific s_0 independent of z, then an

indirect estimate of z_0 via the estimated $\hat{f}_0$ is possible. We have—

$$z_0 = m_0{}'f_0 + s_0 \quad . \quad . \quad . \quad (38)$$

where $m_0{}'$ is a row of occupation loadings for the common factors f_0 of z, and also—

$$\hat{f}_0 = M_0{}'R^{-1}z$$

Substitution in (38), assuming an average $s_0\,(=0)$ gives—

$$\hat{z}_0 = m_0{}'M_0{}'R^{-1}z \quad . \quad . \quad . \quad (39)$$

But—

$$m_0{}'M_0{}' = r_0{}' \quad . \quad . \quad . \quad . \quad (40)$$

and (39) is identical with (37) (Thomson, 1936*a*). If, however, s_0 is not independent of the specifics s of the battery, (40) will not hold, and the estimate (39) made via an estimation of the factors will not agree with the correct estimate (37).

12. *Computation methods.*—The " Doolittle " method of computing regression coefficients is widely used in America (Holzinger, 1937, 32). Aitken's method, used and explained in the text, is in the present author's opinion superior (Aitken, 1937*a* and *b*, with earlier references). Regression calculations and many others are all special cases of the evaluation of a triple matrix product $XY^{-1}Z$, where Y is square and non-singular, and X and Z may be rectangular. The Aitken method writes these matrices down in the form—

$$\begin{array}{c|c} Y & -Z \\ & \\ X & . \end{array}$$

and applies pivotal condensation until all entries to the left of the vertical line are cleared off. All pivots must originate from elements of Y. By giving X and Y special values (including the unit matrix I) the most varied operations can be brought under the one scheme (see Chapter VIII, Section 7).

13. *Bartlett's estimates of factors.*—We have $z = M_0f_0 + M_1f_1$, where f_0 and f_1 are column vectors of the common and specific factors respectively and M_1 is a diagonal

matrix. Bartlett now makes the estimates $\check{f}_0$ such as will minimize the sum of the squares of each person's specifics over the battery of tests, i.e.—

$$\frac{\delta}{\delta f_0}(f_1'f_1) = 0$$

or—

$$\left(\frac{\delta f_1}{\delta f_0}\right)' f_1 = 0$$

i.e.—

$$\begin{aligned}(-M_1^{-1}M_0)'\,(M_1^{-1}z - M_1^{-1}M_0f_0) &= 0 \\ M_0'M_1^{-2}z &= M_0'M_1^{-2}M_0f_0 \\ &= Jf_0,\ \text{say,} \\ \check{f}_0 &= J^{-1}M_0'M_1^{-2}z \quad . \quad (41)\end{aligned}$$

(Bartlett, 1937, 100.)

One could also find the estimated specifics as—

$$\check{f}_1 = (I - M_1^{-1}M_0J^{-1}M_0'M_1^{-1})M_1^{-1}z \quad . \quad . \quad (42)$$

Substituting—

$$z = [M_0 \quad M_1] \begin{bmatrix} f_0 \\ f_1 \end{bmatrix}$$

we get for the relation between $\check{f}$ and f—

$$\begin{bmatrix} \check{f}_0 \\ \check{f}_1 \end{bmatrix} = \left[\begin{array}{c:c} I & J^{-1}M_0'M_1^{-1} \\ \hdashline \cdot & I - M_1^{-1}M_0J^{-1}M_0'M_1^{-} \end{array}\right] \begin{bmatrix} f_0 \\ f_1 \end{bmatrix} = Af \quad (43)$$

and for the covariances of $\check{f}$ we get—

$$AA' = \begin{bmatrix} I + J & \cdot \\ \cdot & I - M_1^{-1}M_0J^{-1}M_0'M_1^{-1} \end{bmatrix} \quad . \quad (44)$$

The error variances and covariances of the common factors are—

$$\begin{aligned}(\check{f}_0 - f_0)(\check{f}_0 - f_0)' &= J^{-1}M_0'M_1^{-1}\,(f_1f_1')\,M_1^{-1}M_0J^{-1} \\ &= J^{-1}M_0'M_1^{-2}M_0J^{-1} = J^{-1} \quad . \quad (45)\end{aligned}$$

(Bartlett, 1937, 100.)

When there is only one common factor, J becomes the familiar quantity—

$$J = S = \Sigma \frac{r_{ig}^2}{1 - r_{ig}^2}$$

(Bartlett, 1935, 200.)

As was first noted by Ledermann *—

$$I + J^{-1} = (M_0'R^{-1}M_0)^{-1} = K^{-1} \quad . \quad (46)$$

(quoted by Thomson, 1938*a*); and using this we see that the back estimates of the original scores from the *regression* estimates $\hat{f}_0$ are identical with the insertion of Bartlett's estimates $\check{f}_0$ in the common-factor part of the specification equations, viz.—

$$M_0K^{-1}M_0'R^{-1}z = M_0J^{-1}M_0'M_1^{-2}z \quad . \quad . \quad (47)$$

(Thomson, 1938*a*.)

Bartlett has pointed out that, using the same identity, in the form $K = J(I - K)$, it is easy to establish the reversible relation between his estimates and regression estimates—

$$\hat{f}_0 = K\check{f}_0, \quad \check{f}_0 = K^{-1}\hat{f}_0 \quad . \quad . \quad (48)$$

(Bartlett, 1938.)

and he summarizes their different interpretation and properties by the formulæ—

$$E\{\hat{f}_0\} = E\{f_0\} = 0, \quad E\{(\hat{f}_0 - f_0)(\hat{f}_0 - f_0)'\} = I - K \quad (49)$$

$$E_1\{\check{f}_0\} = f_0, \quad E_1\{(\check{f}_0 - f_0)(\check{f}_0 - f_0)'\} = J^{-1} = K^{-1}(I - K) \quad . \quad (50)$$

where E denotes averaging over all persons, E_1 over all possible sets of tests (comparable with the given set in regard to the amount of information on the group factors f_0).

14. *Indeterminacy.*—The fact that estimated factors, if the factors outnumber the tests, necessarily have less than unit variance has sometimes been expressed in the case of one common factor by postulating an indeterminate vector i whose variance completes unity. This i may be regarded as the usual error of estimation, and is a function of the specific abilities (Thomson, 1934*b*). The fact that K in Equation (36) is of rank less than its order also expresses the indeterminacy, and allows the factors to be rotated to different positions which nevertheless fulfil all the required conditions. In the hierarchical case the transformation which effects this is (Thomson, 1935*a*)—

$$f = \bar{B}\varphi \quad . \quad . \quad . \quad . \quad . \quad (51)$$

* Letter of October 23, 1937, to Thomson.

where $\acute{B}$ means the required number of rows of—

$$B = I - 2qq'/q'q \quad . \quad . \quad . \quad (52)$$

in which—

$$q_i = l_i/m_i \text{ (see Equation 7)} \quad . \quad (53)$$

as far as there exist tests, after which q is arbitrary.

For—

$$z = Mf = M\acute{B}\varphi = M\varphi$$

since—

$$M\acute{B} = M \quad . \quad . \quad . \quad . \quad . \quad (54)$$

and z is thus expressed by identical specification equations in terms of new factors φ. For such transformations in the case of multiple factors see Thomson, 1936*a*, 140; and Ledermann (paper not yet published).

Indeterminacy is entirely due to the excess of factors over tests, i.e. to the fact that the matrix of loadings M in—

$$z = Mf$$

is not square. It can be in theory abolished by adding a new test which contains no new factor, not even a new specific; or a set of new tests a, b, c . . . which add fewer factors than their number, so that M becomes square (Thomson, 1934*c*; 1935*a*, 258). In the case of a hierarchy each of these tests singly will conform to the hierarchy, so that their saturations l can be found; but jointly they break the hierarchy. If they add no new factors, we shall have—

$$\begin{vmatrix} 1 & l_a & l_b & l_c & . \\ l_a & 1 & r_{ab} & r_{ac} & . \\ l_b & r_{ab} & 1 & r_{bc} & . \\ l_c & r_{ac} & r_{bc} & 1 & . \\ . & . & . & . & . \end{vmatrix} = 0$$

and g can then be found without any indeterminacy from the same equation if we replace the top row of the determinant by the row—

$$g \quad z_a \quad z_b \quad z_c \quad .$$

15. *Finding* g *saturations from an imperfectly hierarchical battery.*—The Spearman formula given in Chapter IX,

Section 5, is the most usual method. A discussion of other methods will be found in Burt, 1936, 283–7. See also Thomson, 1934*a*, 370, for an iterative process modified from Hotelling.

16. *Sampling errors of tetrad-differences.*—The formulæ (16) and (16A) given in the text are both approximations, but appear to be very good approximations. The primary papers are Spearman and Holzinger, 1924 and 1925. Critical examination of the formulæ have been made by Pearson and Moul (1927), and Pearson, Jefferey, and Elderton (1929). Wishart (1928) has considered a quantity P which is equal to $P'N^2/(N-1)(N-2)$, where P' is the tetrad-difference of the covariances a instead of the correlations, and obtained an *exact* expression for the standard deviation σ of P—

$$(N-2)\sigma^2 = \frac{N+1}{N-1} D_{12}D_{34} - D + 3D_{13}D_{31} \qquad (55)$$

where the D's are determinants of the following matrix and its quadrants:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

But approximate assumptions are necessary when the standard deviation of the ordinary tetrad-difference of the correlations is deduced from that of P. The result for the variance of the tetrad-difference is—

$$\frac{N+1}{(N-1)(N-2)}(1-r_{12}{}^2)(1-r_{34}{}^2) - R \quad . \qquad (56)$$

where R is the 4×4 determinant of the correlations.

17. *Selection from a multivariate normal population.*—The primary papers are those of Karl Pearson (1902 and 1912). The matrix form given in the text (Chapter XII, Section 2) is due to Aitken (1934), who employed Soper's device of the moment-generating function, and made a

free use of the notation and methods of matrices. A variant of it which is sometimes useful has been given by Ledermann (Thomson and Ledermann, 1938) as follows. If the original matrix is subdivided in any symmetrical manner:

$$\begin{bmatrix} R_{pp} & R_{pq} & R_{ps} & R_{pt} & . \\ R_{qp} & R_{qq} & R_{qs} & R_{qt} & . \\ R_{sp} & R_{sq} & R_{ss} & R_{st} & . \\ R_{tp} & R_{tq} & R_{ts} & R_{tt} & . \\ . & . & . & . & . \end{bmatrix}$$

and R_{pp} is changed by selection to V_{pp}, then each resulting sub-matrix, including V_{pp} itself, is given by the formula—

$$\left.\begin{aligned} V_{\alpha\beta} &= R_{\alpha\beta} - R_{\alpha p}E_{pp}R_{p\beta} \\ \text{where—} \quad E_{pp} &= R_{pp}^{-1} - R_{pp}^{-1}V_{pp}R_{pp}^{-1} \end{aligned}\right\} \quad . \quad (57)$$

18. *Reciprocity of loadings and factors in persons and traits* (Burt, 1937*b*).—Let W be a matrix of scores centred both by rows and columns. Its dimensions are traits $\times$ persons $(t\,.\,p)$, and its rank is r where r is smaller than both t and p in consequence of the double centring. The two matrices of covariances are WW' for traits and $W'W$ for persons, and by a theorem first enunciated by Sylvester in 1883 (independently discovered by Burt), their non-zero latent roots are the same. If their dimensions differ, i.e. $t \neq p$, the larger one will have additional zero roots. Let the non-zero roots form the diagonal matrix D. Then the principal axes analyses are:

$$W = H_1 D^{\frac{1}{2}} F_1, \text{ dimensions } (t\,.\,r)(r\,.\,r)(r\,.\,p)$$

$$\text{and } W' = H_2 D^{\frac{1}{2}} F_2, \text{ dimensions } (p\,.\,r)(r\,.\,r)(r\,.\,t)$$

where H_1 and H_2 are the latent vectors of WW' and $W'W$, while F_1 is the matrix of factors possessed by persons, F_2 that of factors possessed by traits. From the analysis of W we have, taking the transpose—

$$W' = F_1' D^{\frac{1}{2}} H_1', \text{ dimensions } (p\,.\,r)(r\,.\,r)(r\,.\,t)$$

and comparison of this with the former expression for W' makes the reciprocity of H_2 and F_1', F_2 and H_1', evident.

19. *Oblique factors. Structure and pattern.*—In the specification equations—

$$z = Mf \quad . \quad . \quad . \quad . \quad (3)$$

the matrix of loadings M is called the " pattern," whether the factors are orthogonal or oblique. In the former case M is also the " structure," which is the matrix of correlations of the factors with the tests. " Structure " can also be used in a wider sense to include all the intercorrelations of tests and factors (see Chapters XI and XII for examples). We have, in the narrower sense—

$$\begin{aligned} \text{Structure} &= zf' \\ &= (Mf)f' \\ &= M(ff') \end{aligned}$$

and this is identical with the pattern M if $ff' = I$ (orthogonal factors). Otherwise, with oblique factors,

$$\begin{aligned} \text{Structure} &= \text{pattern} \times ff' \\ &= \text{pattern} \times \text{matrix of factor inter-} \\ &\qquad \text{correlations.} \end{aligned}$$

20. *Boundary conditions.*—These refer to the conditions under which a matrix of correlation coefficients can be explained by factors of limited extent which run each through only a given number of tests. The problem was first raised by Thomson (1919*b*) and a beginning made with its solution (J. R. Thompson, Appendix to Thomson's paper). Various papers by J. R. Thompson culminated in that of 1929, and see also Black, 1929. Thomson returned to the problem in connexion with rotations in the common-factor space (Thomson, 1936*b*), and Ledermann gave rigorous proofs of the theorems enunciated by Thomson and Thompson and extended them (Ledermann, 1936). A necessary condition can now be simply stated, that if the largest latent root of the matrix of correlations exceeds the integer s, then s-factors (that is, factors which run through s tests only and have zero loadings in the other tests) are certainly inadequate. This rule has not been proved to be sufficient as well as necessary, and when applied to the common-factor space only it is certainly not sufficient, though it seems to be a good guide. Ledermann

(1936, 170–4) has given a stringent condition for the case of the common factors, as follows. If we define the nullity of a square matrix as—

$$\text{nullity} = \text{order minus rank}$$

then if it is to be possible to factorize a correlational matrix R of rank r in such a way that the matrix of loadings contains at least r zeros in each of its columns, the sum of the nullities of all the r-rowed principal minors of R must at least be equal to r.

21. *The sampling of bonds.*—The root idea is that of the *complete* family of variates that can be made by all possible additive combinations of bonds from a given pool, and the *complete* family of correlation coefficients between pairs of these. Thomson (1927*b*) mooted the idea and worked out the example quoted in Chapter XVIII. He had earlier (1927*a*) showed that with all-or-none bonds the *most probable* value of a correlation coefficient is $\sqrt{(p_1p_2)}$, where the p's are fractions of the whole pool forming the variates, and the *most probable* value of a tetrad-difference F, zero. Mackie (1928*a*) showed that the *mean* tetrad-difference is zero, and its variance, for F_1—

$$\begin{aligned}\sigma_F{}^2 = \frac{1}{N-1}\Big\{&p_1p_3 + p_2p_4 + p_1p_4 + p_2p_3 - 2(p_1p_2p_3 \\ &+ p_1p_2p_4 + p_1p_3p_4 + p_2p_3p_4) + 4p_1p_2p_3p_4 \\ &+ \frac{2(N-2)}{(N-1)^2}(1-p_1)(1-p_2)(1-p_3) \\ &(1-p_4)\Big\}\end{aligned}$$

where N is the number of bonds in the whole pool. He found for the *mean* value of r_{12} the value $\sqrt{(p_1p_2)}$, and for its variance—

$$\sigma_{r_{12}}{}^2 = \frac{(1-p_1)(1-p_2)}{N-1}$$

This is not the variance of all possible correlation coefficients, but of those formed by taking fractions p_1 and p_2 of the pool. The whole family of correlation coefficients will be widely scattered by reason of the different values of p, "rich" tests having high correlations, and those with low p, low correlations. Mackie (1929) next extended

these formulæ to variable coefficients (i.e. bonds which no longer were all-or-none). He again found the mean value of F to be zero, and for its variance—

$$\sigma_F^2 = \frac{4(N-1)(N-2)}{N^2}\left\{\frac{2}{\pi}\left(1-\frac{2}{\pi}\right)\right\}^2 + \frac{2(N-1)}{N^2}\left\{1-\left(\frac{2}{\pi}\right)^2\right\}^2$$

The presence of $\frac{2}{\pi}$ in this is due to Mackie's limitation to positive loadings of the bonds. Thomson (1935*b*, 72) removed this limitation and found—

$$\sigma_F^2 = \frac{2(N-1)}{N^2}$$

Similarly, Mackie found for variable positive loadings (1929)—

$$\sigma_r^2 = \frac{1}{N}\left\{1-\left(\frac{2}{\pi}\right)^2\right\}$$

and for all loadings Thomson found (1935*b*)—

$$\sigma_r^2 = \frac{1}{N}$$

Thomson suggested without proof that in general, when limits are set to the variability of the loadings of the bonds, resulting in a family of correlation coefficients averaging $\bar{r}$, these correlations will form a distribution with variance—

$$\sigma_r^2 = \frac{1}{N}(1-\bar{r}^2)$$

and will give tetrad-differences averaging zero with a variance—

$$\sigma_F^2 = \frac{4(N-1)(N-2)}{N^2}\left\{\bar{r}(1-\bar{r})\right\}^2 + \frac{2(N-1)}{N^2}(1-\bar{r}^2)^2$$

Summing up, Thomson says (1935*b*, 77–8): "The sampling principle taken alone gives correlations of all values . . . and zero tetrad-differences if N be large. Fitting the sampled elements with weights . . . if the weights may be *any* weights . . . destroys correlation when N is infinite. This means that on the Sampling Theory a certain approxi-

mation to 'all-or-none-ness' is a necessary assumption —not to explain zero tetrad-differences, but to explain the existence of correlations of . . . large size. . . . The most important point in all this appears to me to be the fact that *on all these hypotheses the tetrad-differences tend to vanish.* This tendency appears to be a natural one among correlation coefficients."

A tendency for tetrad-differences to vanish means, of course, a still stronger tendency for large minors of the correlational matrix to vanish. In more general terms, therefore, Thomson's theorem is that in a *complete* family of correlation coefficients the rank of the correlation matrix tends towards unity, and that a *random* sample of variates from this family will (in less strong measure) show the same tendency.

REFERENCES

THIS list is not a bibliography, and makes no pretensions to completeness. It has, on the contrary, been kept as short as possible, and in any case contains hardly any mention of experimental articles. Other references will be found in the works here listed.

References to this list in the text are given thus (Mackie, 1929, 17), or, where more than one article by the same author comes in the same year (Burt, 1937*b*, 84). Throughout the text, however, the two important books by Spearman and by Thurstone are referred to by the short titles *Abilities* and *Vectors* respectively.

The following special abbreviations of the titles of frequently occurring journals are used:

A.J.P. = American Journal of Psychology.
B.J.P. = British Journal of Psychology, General Section.
B.J.E.P. = British Journal of Educational Psychology.
J.E.P. = Journal of Educational Psychology.

AITKEN, A. C., 1934, "Note on Selection from a Multi-variate Normal Population," *Proc. Edinburgh Math. Soc.*, **4**, 106–10.
1937*a*, "The Evaluation of a Certain Triple-product Matrix," *Proc. Roy. Soc. Edinburgh*, **57**, 172–81.
1937*b*, "The Evaluation of the Latent Roots and Vectors of a Matrix," *ibid.*, **57**, 269–304.

ALEXANDER, W. P., 1935, "Intelligence, Concrete and Abstract," *B.J.P. Monograph Supplement* **19**.

ANASTASI, Anne, 1936, "The Influence of Specific Experience upon Mental Organization," *Genetic Psychol. Monographs* **8**, 245–355.
and Garrett (see under Garrett).

BAILES, S., and Thomson (see under Thomson).

BARTLETT, M. S., 1935, "The Statistical Estimation of *G*," *B.J.P.*, **26**, 199–206.
1937*a*, "The Statistical Conception of Mental Factors," *ibid.*, **28**, 97–104.
1937*b*, "The Development of Correlations among Genetic Components of Ability," *Annals of Eugenics*, **7**, 299–302.
1938, "Methods of estimating Mental Factors," *Nature*, **141**, 609–10.

BLACK, T. P., 1929, "The Probable Error of Some Boundary Conditions in diagnosing the Presence of Group and General Factors," *Proc. Roy. Soc. Edinburgh*, **49**, 72–7.

BROWN, W., 1910, "The Correlation of Mental Abilities," *B.J.P.*, **3**, 296–322.

BROWN, W., and Stephenson, W., 1933, "A Test of the Theory of Two Factors," *ibid.*, **23**, 352–70.
1911, and with Thomson, G. H., 1921 and 1925, *The Essentials of Mental Measurement* (Cambridge).
BURT, C., 1917, *The Distribution and Relations of Educational Abilities* (London).
1936, "The Analysis of Examination Marks," a memorandum in *The Marks of Examiners* (London), by Hartog and Rhodes.
1937*a*, "Methods of Factor Analysis with and without Successive Approximation," *B.J.E.P.*, **7**, 172–95.
1937*b*, "Correlations between Persons," *B.J.P.*, **28**, 59–96.
1938, "The Analysis of Temperament," B. J. Medical P., **17**, 158–88.
DAVEY, Constance M., 1926, "A Comparison of Group, Verbal, and Pictorial Intelligence Tests," *B.J.P.*, **17**, 27–48.
DODD, S. C., 1928, "The Theory of Factors," *Psychol. Rev.*, **35**, 211–34 and 261–79.
1929, "The Sampling Theory of Intelligence," *B.J.P.*, **19**, 306–27.
ELDERTON, E. M., Jeffery and Pearson (see under Pearson).
EMMETT, W. G., 1936, "Sampling Error and the Two-factor Theory," *B.J.P.*, **26**, 362–87.
ETHERINGTON, I. M. H., 1932, "On Errors in Determinants," *Proc. Edinburgh Math. Soc.*, **3**, 107–17.
FILON, L. N., and Pearson (see under Pearson).
GARNETT, J. C. M., 1919*a*, "On Certain Independent Factors in Mental Measurement," *Proc. Roy. Soc.*, A, **96**, 91–111.
1919*b*, "General Ability, Cleverness, and Purpose," *B.J.P.*, **9**, 345–66.
GARRETT, H. E., and Anastasi, Anne, 1932, "The Tetrad-difference Criterion and the Measurement of Mental Traits," *Annals New York Acad. Sciences*, **33**, 233–82.
HARMAN, H. H., and Holzinger (see under Holzinger).
HART, B., and Spearman (see under Spearman).
HEYWOOD, H. B., 1931, "On Finite Sequences of Real Numbers," *Proc. Roy. Soc.*, A, **134**, 486–501.
HOLZINGER, K. J., 1935, *Preliminary Reports on Spearman-Holzinger Unitary Trait Study*, No. 5; *Introduction to Bifactor Theory* (Chicago).
1937*a*, *Student Manual of Factor Analysis* (Chicago).
and Harman, H. H., 1937*b*, "Relationships between Factors obtained from Certain Analyses," *J.E.P.*, **28**, 321–45.
and Spearman (see under Spearman).
Comparison of Two Factorial Analyses, *Psychometrika*, **3**, 45–60.
HOTELLING, H., 1933, "Analysis of a Complex of Statistical Variables into Principal Components," *J.E.P.*, **24**, 417–41 and 498–520.
1935*a*, "The Most Predictable Criterion," *J.E.P.*, **26**, 139–42.

HOTELLING, H., 1935*b*, "Simplified Calculation of Principal Components," *Psychometrika*, **1**, 27–35.

—— 1936, "Relations between Two Sets of Variates," *Biometrika*, **28**, 321–77.

IRWIN, J. O., 1932, "On the Uniqueness of the Factor *g* for General Intelligence," *B.J.P.*, **22**, 359–63.

1933, "A Critical Discussion of the Single-factor Theory," *ibid.*, **23**, 371–81.

JEFFERY, G. B., Pearson, and Elderton (see under Pearson).

KELLEY, T. L., 1923, *Statistical Method* (New York).

1928, *Crossroads in the Mind of Man* (Stanford and Oxford).

1935, *Essential Traits of Mental Life* (Harvard).

LEDERMANN, W., 1936, "Some Mathematical Remarks concerning Boundary Conditions in the Factorial Analysis of Ability," *Psychometrika*, **1**, 165–74.

1937, "On the Rank of the Reduced Correlational Matrix in Multiple-factor Analysis," *ibid.*, **2**, 85–93.

1938*a*, "A Shortened Method of Estimation of Mental Factors by Regression," *Nature*, **141**, 650.

1938*b*, "Note on Professor Godfrey Thomson's Article on the Influence of Univariate Selection on Factorial Analysis," *B.J.P.*, **29**, 69–73.

and Thomson (see under Thomson).

MACKIE, J., 1928*a*, "The Probable Value of the Tetrad-difference on the Sampling Theory," *B.J.P.*, **19**, 65–76.

1928*b*, "The Sampling Theory as a Variant of the Two-factor Theory," *J.E.P.*, **19**, 614–21.

1929, "Mathematical Consequences of Certain Theories of Mental Ability," *Proc. Roy. Soc. Edinburgh*, **49**, 16–37.

MOUL, M., and Pearson (see under Pearson).

PEARSON, K., 1902, "On the Influence of Natural Selection on the Variability and Correlation of Organs," *Phil. Trans. Roy. Soc. London*, **200**, A, 1–66.

1912, "On the General Theory of the Influence of Selection on Correlation and Variation," *Biometrika*, **8**, 437–43.

and Filon, L. N. G., 1898, "On the Probable Errors of Frequency Constants and on the Influence of Random Selection on Variation and Correlation," *Phil. Trans. Roy. Soc. London*, **191**, A, 229–311.

Jeffery, G. B., and Elderton, E. M., 1929, "On the Distribution of the First Product-moment Coefficient, etc.," *Biometrika*, **21**, 191–2.

and Moul, M., 1927, "The Sampling Errors in the Theory of a Generalized Factor," *ibid.*, **19**, 246–91.

PIAGGIO, H. T. H., 1933, "Three Sets of Conditions Necessary for the Existence of a *g* that is Real and Unique except in Sign," *B.J.P.*, **24**, 88–105.

Price, B., 1936, "Homogamy and the Intercorrelation of Capacity Traits," *Annals of Eugenics*, **7**, 22–7.

Spearman, C., 1904, "General Intelligence objectively Determined and Measured," *A.J.P.*, **15**, 201–93.

1913, "Correlation of Sums or Differences," *B.J.P.*, **5**, 417–26.

1914, "The Theory of Two Factors," *Psychol. Rev.*, **21**, 101–15.

1926 and 1932, *The Abilities of Man* (London).

1928, "The Substructure of the Mind," *B.J.P.*, **18**, 249–61.

1931, "Sampling Error of Tetrad-differences," *J.E.P.*, **22**, 388.

and Hart, B., 1912, "General Ability, its Existence and Nature," *B.J.P.*, **5**, 51–84.

and Holzinger, K. J., 1924, "The Sampling Error in the Theory of Two Factors," *B.J.P.*, **15**, 17–19.

and Holzinger, K. J., 1925, "Note on the Sampling Error of Tetrad-differences," *ibid.*, **16**, 86–8.

and Holzinger, K. J., 1929, "Average Value for the Probable Error of Tetrad-differences," *ibid.*, **20**, 368–70.

Stephenson, W., 1931, "Tetrad-differences for Verbal Sub-tests relative to Non-verbal Sub-tests," *J.E.P.*, **22**, 334–50.

1935*a*, "The Technique of Factor Analysis," *Nature*, **136**, 297.

1935*b*, "Correlating Persons instead of Tests," *Character and Personality*, **4**, 17–24.

1936*a*, "A new Application of Correlation to Averages," *B.J.E.P.*, **6**, 43–57.

1936*b*, "The Inverted-factor Technique," *B.J.P.*, **26**, 344–61.

1936*c*, "Introduction to Inverted-factor Analysis, with some Applications to Studies in Orexis," *J.E.P.*, **27**, 353–67.

1936*d*, "The Foundations of Psychometry: Four Factor Systems," *Psychometrika*, **1**, 195–209.

and Brown (see under Brown).

Thompson, J. R. (see under Thomson, 1919*b*).

1929, "The General Expression for Boundary Conditions and the Limits of Correlation," *Proc Roy. Soc. Edinburgh*, **49**, 65–71.

Thomson, G. H., 1916, "A Hierarchy without a General Factor," *B.J.P.*, **8**, 271–81.

1919*a*, "On the Cause of Hierarchical Order among Correlation Coefficients," *Proc. Roy. Soc.*, A, **95**, 400–8.

1919*b*, "The Proof or Disproof of the Existence of General Ability" (with Appendix by J. R. Thompson), *B.J.P.*, **9**, 321–36.

1927*a*, "The Tetrad-difference Criterion, *B.J.P.*, **17**, 235–55.

1927*b*, "A Worked-out Example of the Possible Linkages of Four Correlated Variables on the Sampling Theory," *ibid.*, **18**, 68–76.

1934*a*, "Hotelling's Method modified to give Spearman's *g*," *J.E.P.*, **25**, 366–74.

1934*b* "The Meaning of *i* in the Estimate of *g*," *B.J.P.*, **25**, 92–9.

THOMSON, G. H., 1934*c*, "On measuring *g* and *s* by Tests which break the *g*-hierarchy," *ibid.*, **25**, 204–10.

1935*a*, "The Definition and Measurement of *g* (General Intelligence)," *J.E.P.*, **26**, 241–62.

1935*b*, "On Complete Families of Correlation Coefficients and their Tendency to Zero Tetrad-differences: including a Statement of the Sampling Theory of Abilities," *B.J.P.*, **26**, 63–92.

1935*c*, "The Factorial Analysis of Human Abilities," *The Human Factor*, **9**, 180–5. "A Reply," by R. H. Thouless, and "A Rejoinder," by Thomson, *ibid.*, 358–63.

1936*a*, "Some Points of Mathematical Technique in the Factorial Analysis of Ability," *J.E.P.*, **27**, 37–54.

1936*b*, "Boundary Conditions in the Common-factor Space, in the Factorial Analysis of Ability," *Psychometrika*, **1**, 155–63.

1937, "Selection and Mental Factors," *Nature*, **140**, 934.

1938*a*, "Methods of estimating Mental Factors," *ibid.*, **141**, 246.

1938*b*, "The Influence of Univariate Selection on the Factorial Analysis of Ability," *B.J.P.*, **28**, 451–9.

1938*c*, "On Maximizing the Specific Factors in the Analysis of Ability," *B.J.E.P.* (in the press).

and Bailes, S., 1926, "The Reliability of Essay Marks," *The Forum of Education*, **4**, 85–91.

and Brown, W. (see under Brown).

and Ledermann, W., 1938, "The Influence of Multi-variate Selection on the Factorial Analysis of Ability," *B.J.P.* (in the press).

THORNDIKE, E. L., 1925, *The Measurement of Intelligence* (New York).

THOULESS, R. H., 1935 (see under Thomson, 1935*c*).

THURSTONE, L. L., 1933, *A Simplified Multiple-factor Method* (Chicago).

1935, *The Vectors of Mind* (Chicago).

1938, "Primary Mental Abilities," *Psychometric Monograph No.* 1 (Chicago).

TRYON, R. C., 1932*a*, "Multiple Factors *vs* Two Factors as Determiners of Abilities," *Psychol. Rev.*, **39**, 324–51.

1932*b*, "So-called Group Factors as Determiners of Ability," *ibid.*, **39**, 403–39.

1935, "A Theory of Psychological Components—an Alternative to Mathematical Factors," *ibid.*, **42**, 425–54.

WISHART, J., 1928, "Sampling Errors in the Theory of Two Factors," *B.J.P.*, **19**, 180–7.

WILSON, E. B., 1928*a*, "On Hierarchical Correlation Systems," *Proc. Nat. Acad. Sc.*, **14**, 283–91.

1928*b*, "Review of *The Abilities of Man*," *Science*, **67**, 244–8.

1933*a*, "On the Invariance of General Intelligence," *Proc. Nat. Acad. Sc.*, **19**, 768–72.

WILSON, E. B., 1933*b*, "Transformations preserving the Tetrad Equations," *ibid.*, **19**, 882–4.

1933*c*, "On Overlap," *ibid.*, **19**, 1039–44.

1934, "On Resolution into Generals and Specifics," *ibid.*, **20**, 193–6.

and Worcester, Jane, 1934, "The Resolution of Four Tests," *ibid.*, **20**, 189–92.

WORCESTER, Jane, and Wilson (see under Wilson).

WRIGHT, S., 1921, "Systems of Mating," *Genetics*, **6**, 111–78.

INDEX

The numbers refer to pages. Terms which occur repeatedly through the book are only indexed on their first occurrence or where they are defined. Most chapter and section headings are indexed. The names Spearman and Thurstone (which occur so frequently) are not indexed, nor is the author's name.